

University of South Wales



2059310



105 Cathays Terrace, Cardiff CF24 4HU  
South Wales, U.K. Tel: (029) 2039 5882  
[www.bookbindersuk.com](http://www.bookbindersuk.com)  
Email: [mail@bookbindersuk.com](mailto:mail@bookbindersuk.com)

University of Glamorgan  
Prifysgol Morgannwg



# Computational techniques for the geo-temporal analysis of crime and disorder data

Jonathan Corcoran

School of Computing  
University of Glamorgan  
Pontypridd

A thesis submitted in partial fulfilment of the requirements of the  
University of Glamorgan for the degree of Doctor of Philosophy

# TABLE OF CONTENTS

<b>TABLE OF CONTENTS .....</b>	<b>(i-iii)</b>
<b>LIST OF FIGURES .....</b>	<b>(iv)</b>
<b>LIST OF TABLES .....</b>	<b>(v)</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>(vi)</b>
<b>DECLARATIONS.....</b>	<b>(vii-viii)</b>
<b>ABSTRACT .....</b>	<b>(ix)</b>
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 Background.....	1
1.2 The geography of crime and disorder.....	1
1.2.1 The current toolkit - analysing crime and disorder data.....	2
1.2.2 The future toolkit - predicting crime and disorder data.....	2
1.3 Objectives of research .....	3
1.4 Study region.....	4
1.5 Outline and content of thesis .....	4
<b>2 LITERATURE REVIEW .....</b>	<b>6</b>
2.1 Background.....	6
2.2 Criminological Theory .....	7
2.3 Crime mapping and crime analysis .....	10
2.3.1 Crime mapping systems .....	10
2.3.2 Crime analysis systems .....	10
2.4 Geocomputational techniques.....	11
2.4.1 Spatial distribution .....	11
2.4.2 Predictive techniques.....	13
2.5 Future trends and research focus .....	14
2.6 Summary.....	16
<b>3. COMPUTERISED MAPPING AND SPATIAL ANALYSIS AS A STRATEGIC AND OPERATIONAL TOOL .....</b>	<b>17</b>
3.1 Introduction.....	17
3.2 Mapping Systems in the ES .....	18
3.2.1 Police.....	18
3.2.2 Fire .....	19
3.2.3 Ambulance .....	20
3.3 National surveys.....	20
3.3.1 Mapping or No Mapping.....	21
3.3.2 Information output & dissemination .....	22
3.3.3 Software selection .....	23
3.3.4 Mapping and Analysis techniques .....	25
3.3.5 Customisation and Global Positioning Systems (GPS).....	26
3.3.6 Perceived usefulness of computerised mapping.....	27
3.3.7 Implications, problems and looking to the future .....	28
3.4 Summary and Conclusions.....	30

<b>4</b>	<b>ETHICAL DILEMMAS IN CRIME AND DISORDER ANALYSIS: AN OVERVIEW.....</b>	<b>32</b>
4.1	Introduction.....	32
4.2	The ethics of crime analysis and prediction.....	32
4.3	A case study .....	34
4.4	Conflicts between case study proposal and ethical approval .....	35
4.5	Appeasement of the conundrum .....	36
4.6	Summary .....	39
<b>5</b>	<b>MODELLING DISPARATE CRIME AND DISORDER DATA .....</b>	<b>40</b>
5.1	Introduction.....	40
5.2	The HASCADE model.....	42
5.3	Data requirements and issues .....	46
5.3.1	Technical Issues .....	46
5.3.2	Security Issues.....	47
5.3.3	Data issues.....	47
5.4	Towards a new approach.....	48
5.4.1	Spatial and statistical techniques.....	48
5.4.2	Community safety public consultation.....	50
5.4.3	Spatial Analysis.....	51
5.4.4	Statistical Analysis .....	53
5.5	HASCADE results .....	55
5.6	Future developments .....	55
5.7	Summary and Conclusions.....	57
<b>6</b>	<b>EXPLORATION OF INCIDENT AND INJURY DATA.....</b>	<b>59</b>
6.1	Introduction.....	59
6.2	Data procurement.....	59
6.3	Data Inspection .....	60
6.3.1	Erroneous and missing data .....	60
6.3.2	A&E .....	63
6.3.3	Police.....	70
6.3.4	Comparison of Police and A&E data .....	87
6.4	Summary and Conclusions.....	89
<b>7</b>	<b>MODELLING CRIME DATA FOR PREDICTION.....</b>	<b>91</b>
7.1	Introduction.....	91
7.2	Initial attempts at modelling .....	92
7.3	Artificial Neural Networks Models for Prediction.....	93
7.4	The Gamma Test.....	95
7.5	Empirical evidence and output.....	96
7.5.1	Stage One: Point density analysis .....	97
7.5.2	Stage Two: Geographic representation and cluster analysis.....	98
7.5.3	Stage Three: Allocation of centroids to clusters .....	99
7.5.4	Stage Four: Relating incidents to cluster boundaries.....	100
7.6	Application of Gamma test to the cluster data .....	101
7.7	Forecasting using Artificial Neural Networks .....	101
7.7.1	Network topology .....	102
7.7.2	The number of nodes in the hidden layers .....	102
7.7.3	Terminating the training procedure.....	103
7.7.4	Partitioning the vectors into training and test sets.....	103
7.8	Experimental Work .....	104
7.8.1	Comparison between ANN, Linear Regression and Random Walk .....	104
7.9	Discussion of results .....	106
7.10	Summary and Conclusions.....	107



<b>8 CONCLUSIONS AND FUTURE WORK .....</b>	<b>108</b>
8.1 Introduction.....	108
8.2 Concluding remarks .....	112
8.3 Contribution to knowledge.....	112
8.4 Potential avenues for future research .....	113
8.5 Final Remarks .....	118
<b>APPENDIX A – EMERGENCY SERVICE QUESTIONNAIRE (POLICE) .....</b>	<b>A1-A9</b>
<b>APPENDIX B – EMERGENCY SERVICE SURVEY RESULTS .....</b>	<b>B1-B3</b>
<b>APPENDIX C – POLICE SPATIAL DATA DISTRIBUTIONS .....</b>	<b>C1-C4</b>
<b>APPENDIX D – COMMUNITY CONSULTATION QUESTIONNAIRE.....</b>	<b>D1-D4</b>
<b>APPENDIX E – PUBLISHED PAPERS .....</b>	<b>E1-E85</b>
 <b>Paper 1</b> Corcoran, J. and J. A. Ware (2002) “Helping with enquiries.” <u>GeoConnexion</u> . 1(3): 36-40 .....	 E1-E11
 <b>Paper 2</b> Sivarajasingam, V., J. Corcoran, D. Jones, J.A.Ware and J. Shepherd. “Relations between violence, calendar events and ambient conditions.” Accepted for publication in the Journal of Injury Prevention.....	 E12-E22
 <b>Paper 3</b> Corcoran, J. and B. Bowen Thomson “New Insights into Community Safety: Application of the HASCADE Model.” Accepted for publication in the British Journal of Community Justice  A summary of this paper was presented at the Managing the Strategy Conference, Cardiff 2003.....	  E23-E41
 <b>Paper 4</b> Corcoran, J., I.D Wilson and J.A. Ware “Predicting the Geo-Temporal Variations of Crime and Disorder.” Accepted for publication in the International Journal of Forecasting (Special Issue on Crime Forecasting)  A summary of this paper was presented at the 6 <sup>th</sup> Annual Crime Mapping Conference in Denver, USA 2002.  A poster summarising the work was presented at <i>Geovisualisation</i> , Albufira, Portugal 2002. ....	   E42-E62
 <b>Paper 5</b> Corcoran, J. and J.A. Ware “Crime hot spot prediction: a framework for progress.” <u>Socio-economic applications in geographical information science. Innovations in GIS 9</u> . D. B. Kidner, G. Higgs and S. White. London, Taylor & Francis: (2003): 51-67  A summary of this paper was presented at GISRUUK 2001 and appears in the proceedings: Corcoran, J. and J. A. Ware (2001). <u>Data Clustering using Artificial Neural Networks as a Precursor to Crime Hot Spot Prediction</u> . GISRUUK, University of Glamorgan, Wales: 249-253.  A summary of this paper was also presented at the 5 <sup>th</sup> Annual Crime Mapping Conference in Dallas 2001.....	   E63-E75

## LIST OF FIGURES

Figure 3.1 Uptake of mapping technologies timeline .....	21
Figure 3.2 Mapping software selection.....	24
Figure 3.3 Database software selection.....	24
Figure 3.4 Contextual data usage.....	25
Figure 3.5 Perceived usefulness of computerised mapping .....	28
 Figure 5-1 The HASCADE model.....	 43
Figure 5-2 The HASCADE count system.....	44
Figure 5-3 Method of validating the boundary network .....	52
Figure 5-4 Factor map showing PGAs.....	55
 Figure 6-1 Total assault cases by week.....	 63
Figure 6-2 Breakdown of assault by gender .....	64
Figure 6-3 Three monthly totals.....	66
Figure 6-4 Total incidents by month.....	71
Figure 6-5 Volumes of incident (a) monthly, (b) weekly and (c) hourly.....	74
Figure 6-6 Percentage variation of incident by weekday and hour (a) c10 (b) c60 (c) d10 and (d) d90.....	75
Figure 6-7 The snapshot model.....	76
Figure 6-8 (a) Single view port (b) Dual view port animation player.....	77
Figure 6-9 Subset from a typical animated sequence (specifically c10 focusing on the city centre region) .....	78
Figure 6-10 NNA by day (a) c10 (b) c60 (c) d10 and (d) d90 .....	80
Figure 6-11 NNA by hour (a) c10 (b) c60 (c) d10 and (d) d90.....	81
Figure 6-12 Variation in the utilisation distribution by week (90% probability).....	82
Figure 6-13 Variation in the utilisation distribution by hour (90% probability) for (a) c10 (b) c60 (c) d10 and (d) d90 .....	82
Figure 6-14 Flows of c10 incidents within the city centre activity region.....	84
Figure 6-15 (a) Monthly trend (b) Daily trend.....	88
 Figure 7.1 Policing boundaries (sector level) for the study area.....	 92
Figure 7.2 The CLAP/GT/Predictive process model .....	93
Figure 7.3 (a) A diagram of an ANN (b) A diagram of a neuron. ....	94
Figure 7.4 Moving window technique .....	94
Figure 7.5 The Gamma Statistic and the Gradient/slope .....	96
Figure 7.6 Incidence density algorithm.....	97
Figure 7.7 Centroids of higher than average crime incidence .....	98
Figure 7.8 Clusters of high crime incidence .....	98
Figure 7.9 Clusters of high crime incidence, displayed using the CLAP interface .....	99
Figure 7.10 Centroid clustering algorithm.....	99
Figure 7.11 Incident to cluster boundary allocation algorithm.....	100
Figure 7.12 Example of increasing embedding (City Centre, cluster Four) .....	102
Figure 7.13 Gamma Scatter Plot showing the trend line .....	102
Figure 7.14 M-Test for the City Centre .....	104
Figure 7.15 Incidence and forecast of violent crime (Cluster 7).....	105
Figure 7.16 Incidence and forecast of violent crime (City Centre).....	106
 Figure 8-1 Cardiff city centre.....	 115
Figure 8-2 Using a KSOM to partition a dataset.....	118

## LIST OF TABLES

Table 5.1 Factor classification of community vulnerabilities .....	43
Table 5.2 Classification of small areas into factors. ....	45
Table 5.3 Minimum data provision proforma .....	47
Table 5.4 Community vulnerabilities and related data sets .....	50
Table 5.5 An example of the correlation coefficients (significance at 99% denoted by a shaded cells).....	54
Table 6-1 Data sets used in the analysis.....	61
Table 6-2 Breakdown of assault by day .....	64
Table 6-3 Distribution by age and gender .....	65
Table 6-4 Age group by day .....	65
Table 6-5 Difference between seasonal totals.....	67
Table 6-6 Pearson correlation coefficients for temperature, rainfall and sunshine against assault intake (significance level in parentheses).....	67
Table 6-7 Largest A&E attendance's.....	68
Table 6-8 Results of Moran I for each incident type .....	72
Table 7-1 Crime incidence by cluster .....	101
Table 7-2 Results for Cluster Seven .....	105
Table 7-3 Results for Cluster Four.....	106
Table 8-1 Typical errors found in Police database records .....	114

## ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisors, Prof. J.A. Ware, Dr. J.M. Ware and Dr. D.B. Kidner for their continued support and guidance throughout the duration of the research. I wish to extend a special thanks to Prof. J.A. Ware for accompanying me on many scenic miles of international travelling as a direct result of the research.

Thanks also to each of the collaborators, South Wales Police, Heath Hospital Cardiff, and the Crime and Disorder Reduction Partnerships of Cardiff and Barry.

Thanks are due to my fellow researchers for providing an excellent working atmosphere, and finally, to my parents for their continued love and support.

## **CERTIFICATE OF RESEARCH**

This is to certify that, except where specific reference is made, the work presented in this thesis is the result of the investigation undertaken by the candidate.

**CANDIDATE**

.....

**DIRECTOR  
OF  
STUDIES**

.....

## **CERTIFICATE OF RESEARCH**

This is to certify that neither this thesis or any part of it has been presented or is being currently submitted in candidature for any other degree other than the degree of Doctor of Philosophy of the University of Glamorgan.

**CANDIDATE**



.....

## ABSTRACT

### **Computational Techniques for the Geo-Temporal Analysis of Crime and Disorder Data**

This thesis presents research that has resulted in the development of techniques and methodologies through which crime and disorder can be analysed. These are applicable for use by law enforcement agencies and those charged with the management of crime, injury and community safety. A fundamental element in developing the techniques presented in this thesis was an appreciation of those currently in use. National surveys of all three emergency services (Police, Fire and Ambulance) offered an insight into the use and uptake of computerised mapping technologies for both operational and strategic purposes. The surveys highlighted current trends in and attitudes to such technologies in addition to the identification of potential future developments. Further insight was gained through working with local Crime and Disorder Reduction Partnerships. Here a framework was designed and implemented by which crime and disorder data could be audited using geo-statistical techniques.

Two key exploratory techniques are presented in the thesis that develop the currently available tool set. The first, uses animation embedded within a Geographical Information System through which multiple snap shots of criminal activity can be aggregated and played back to the user in the form of an animated sequence. The second, develops the use of animation to provide a way by which observed movements (or flows of criminal activity) can be qualified through analysis of clusters over time. This is achieved by quantifying volume, centroid and direction of movement.

The effective allocation of finite resources is a key issue for each of the public services covered by this thesis. A novel methodology is proposed to address this requirement. The methodology uses a geographical scanning algorithm to generate clusters of sufficient data from which Artificial Neural Networks can be trained to model a simple cause and effect relationship.

# 1 INTRODUCTION

*This Chapter presents the thesis outline together with the objectives and motivations for the research. The Chapter concludes by detailing the content of subsequent Chapters.*

---

## 1.1 Background

Crime is an unfortunate reality of society, affecting either directly or indirectly, the vast majority of the populace. Its control is charged to a range of both nationally and locally governed agencies. Key to achieving their missions is the ability to assimilate an understanding of criminal dynamics, which are inherently complex. Fundamental to gaining this understanding is a correct application of technology to collect and analyse data. Their contribution can provide a platform from which decisions can be formulated to address directly issues of crime and vulnerability. For example, in policing, this may translate into the ability to accurately and consistently record the type and locales of crimes. This may be followed by a series of specific techniques designed to interrogate the recorded information in order to help identify the existence of trends that can offer a contribution towards policing decisions.

## 1.2 The geography of crime and disorder

Crime data and the computational tools that are used for its collection and analysis have, over recent years, grown in importance. Academics and practitioners have seen value in their potential to analyse crime and disorder issues.

A central theme in the geo-analysis of crime and disorder data is the quest to better understand their dynamics, which in turn can be applied to formulate targeted responses. This thesis continues the theme in two main areas.



The first is concerned with improving the current toolbox. The second assesses whether crime can be predicted over both time and space and, if so, which methodologies are most applicable. (Relevant to both questions is the issue of data availability. Issues of data availability and the ethical restrictions of their use are presented in Chapter Four.) The following sections take each of the two areas and discuss the motivations for research.

### **1.2.1 The current toolkit - analysing crime and disorder data**

The ability to automate the generation of user-driven requests, and present the results in an easily comprehensible visualisation, forms the core of many geographical crime analysis tools and methodologies. Typically, these tools and methodologies attempt to embrace questions such as “are there any patterns or trends, and if so, what are they and when and where do they occur?” If the answers to these questions are deemed significant, then preventative measures can then be deployed to the identified locales. The subsequent success or failure of the preventative measures can then be judged on an assessment of any sustained reductions or displacement of crime.

### **1.2.2 The future toolkit - predicting crime and disorder data**

The capability to predict future criminal activity is a very idealistic notion. The facility to confidently respond to the “what”, “when” and “where” questions are of value in the quest to minimise crime, disorder and community vulnerabilities. Utilisation of such intelligence may involve resource allocation (for example, police officers) to an identified region at a specific time to assist in the prevention of criminal activity. Over longer granularities of time, this may involve the instigation of targeted initiatives (for example a lock fitting scheme in areas identified as vulnerable to burglary). The predictive system should also be capable of learning any changes to previous patterning, following initiation of preventative measures. Thus the system would be continually evolving, capable of responding to changes in the environment (social and physical) in which crime is committed.

The first obvious question is whether there is an operational necessity for such a system. There is the possibility that such intelligence already exists through the experiences of police officers. If, for example, a police officer was invited to identify high crime areas, they could through personal experiences be able to identify particular localities and timings as being potentially problematic. However, one possible problem arises when officers are either transferred, promoted or retire and their personal experiences and knowledge are lost. In addition, the fact that such intelligence is based on experience, means that the knowledge of individual officers is neither continuous, consistent, nor formalised, and easily overlooked

when such individuals take no part in policy formulation. It is also noted that recent research that suggested that perceptions of patrolling officers may not be as well defined as initially thought and is dependent upon the crime type (Ratcliffe 2001).

Thus the role of a predictive system would be of value in many situations, gelling together the experiences of individuals and presenting them in a generic format applicable to the operational and strategic decision making process. However, such a predictive capability should not be seen as being in direct competition with the police officer, but more of a complementary tool through which informed opinions can be jointly based. This function can be seen more as a unifying one that combines the specific experiences of officers with the outputs from a predictive system.

Thus far, discussion has focused on the application of a predictive facility as a policing tool. However, crime, disorder and injury are recorded, monitored and administered by a range of additional agencies (for example, local authority departments, probation service and health authorities) that could also benefit from such tools, either individually or collectively.

Given the above motivations for research, the following sections detail its objectives and the study region.

### **1.3 Objectives of research**

Following an ethical appeasement process (presented in greater detail in Chapter Four), research objectives were finalised in the light of both available data and imposed restrictions on their use:

- Evaluate the current use and uptake of computerised mapping technologies for operational and strategic incident mapping;
- Develop a GIS that describes crime and disorder issues, and integrates all available relevant data sets;
- Implement and evaluate exploratory and predictive techniques for use on crime and incident databases;
- Develop a GIS for visualisation of geospatial information retrieved from an analytical engine comprising Artificial Intelligence paradigms.

### **1.4 Study region**

Cardiff (United Kingdom), comprises 28 electoral wards covering 524 acres (46 square miles) and forms the selected study locality for this research. Cardiff, officially the capital of Wales since 1955, is located in the south-eastern corner of the country bordering the Bristol Channel to the south and the M4 motorway to the North. The city houses a population of 327,500 inhabitants (mid year estimate for 2000<sup>1</sup>) and consists of 132,552 households (of which approximately 5,500 are unoccupied) and 77.5% comprise owner occupied properties. The city also houses a large student population (in excess of 25,000 in 1999).

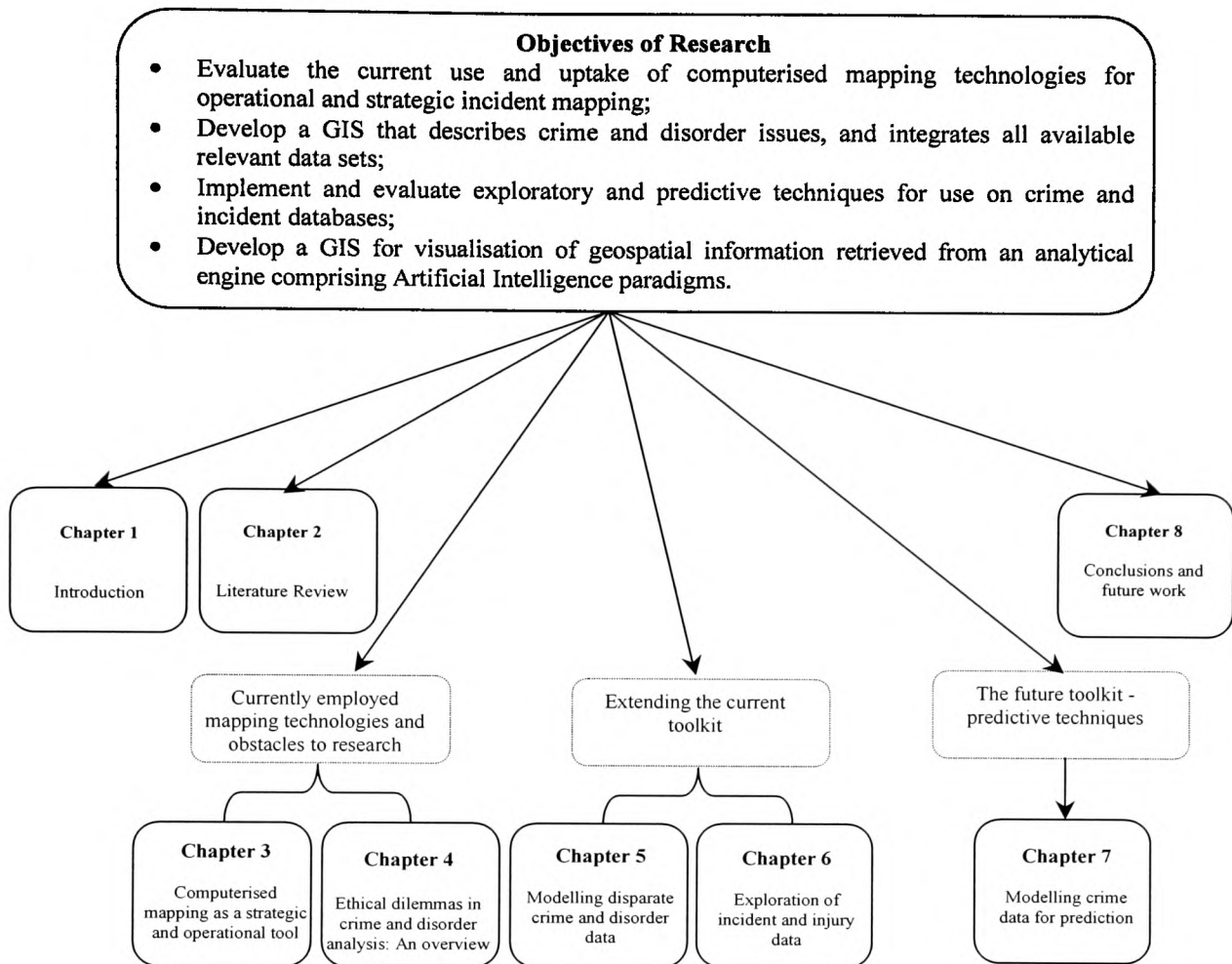
Key to the formation of the Cardiff region was the industrial revolution, which instigated a rapid growth in its size and development as an international trading port. Since the Second World War, however, the decline and ultimate termination of the coal trade, contributed to high unemployment levels during the 1980s. The dockland areas have like many others in Britain experienced decline, although they have subsequently undergone extensive redevelopment, encouraging the establishment of commercial, professional and retail enterprises in the area. Extensive investment within the city centre has seen the rebuilding of a national sporting stadium, with an all-seated capacity of 72,500, which since opening in 1999 has attracted a variety of national and international sporting events, in addition to concerts.

### **1.5 Outline and content of thesis**

The remainder of the thesis is divided into seven Chapters. Chapter Two discusses the relevant issues and developments in crime mapping, analysis and prediction. Chapter Three presents results of a national survey of the three Emergency Services (Police, Fire, and Ambulance) establishing the operational uptake and utilisation of computerised mapping facilities. Chapter Four presents an ethical evaluation that the research underwent at time of inception. Association with local Crime and Disorder Reduction Partnerships, Chapter Five presents a geo-statistical methodology through which formerly disparate community data can be analysed in a holistic manner. Chapter Six presents a series of spatial, temporal, spatio-temporal and statistical techniques for the exploration of crime, injury and contextual data. Following the exploratory process Chapter Seven describes the application of an Artificial Neural Networks (ANN) technique developed to assist in the prediction of future criminal events. Chapter Eight offers an overview of the techniques and methodologies presented in previous Chapters, concluding with a summary of some potential avenues for future research.

---

<sup>1</sup> Statistics taken from Cardiff County Council Research Centre



## 2 LITERATURE REVIEW

*This chapter discusses the relevant literature pertaining to the field of crime and disorder mapping, from its inception through to current and potential avenues of research. Particular emphasis is placed on relevant criminological theories and geocomputational techniques.*

---

### 2.1 Background

The mapping of crime has a long history as a tool for understanding crime's spatial distributions. It can be traced back as far as the 19th century in France (Guerry 1833; Quêtelet 1842), where mapping was first utilised to visualise and analyse crime information. In time this so called "*cartographic school of criminology*" (Sutherland and Cressey 1970; Phillips 1972) spread to Britain (Fletcher 1849; Mayhew 1862) although it dissipated soon after, with resistance to the manual and time consuming cartographic process. At this point, the traditional focus on crime redefined to sociological and psychological based enquiries (Georges 1978). Interest and awareness of a spatial approach in criminology has been well established falling under the generic title of aerial studies (Baldwin 1975). The majority of these studies, however, tended to utilise a spatial framework as a supplement to sociological and physiological thought and issues. Interest and greater focus by geographers began to re-emerge in the late 1970's where the study of crime became the motivation of more geographic enquires (Scott 1972; Harries 1974; Pyle 1974; Herbert 1976; Herbert 1982). The geography of crime, however, received some early criticism, with questions raised as to whether it provided a legitimate methodological approach (Peet 1975) although its use was subsequently defended (Harries 1975; Lee 1975; Phillips 1975).

The notion is that crime occurrence can be modelled in a geographical context and a greater understanding of the patterns of that distribution can be sought.

*"The objectives of the geography of crime are to describe and map the spatial distribution of crime in greater detail and meaning than has been done before. This field of research attempts to relate the spatial patterns of crime to the environmental, social, historical, psychological (cognitive), and economic variables that may explain these patterns. It hopes to develop associational and predictive models that explain crime manifestation in regard to locale. Last but not least, it is hoped that its contribution to the analysis of the dynamics of crime manifestation will help those charged with the responsibility of crime control to assess better the effectiveness of programs they currently use"* (Georges 1978: 20).

## 2.2 Criminological Theory

Crimes are a human phenomena and the assumption is that their occurrence in a spatial and spatio-temporal framework are not random. Identification of spatial patterning has seen the promotion of several theories and has been the subject of many philosophical debates in an attempt to help explain their manifestation. Four theories of relevance to geographical studies of crime are now described. The first three have been termed "*place based theories*" (Anselin, *et al.* 2000: 219) where the objective is to derive an understanding of mechanisms upon individual actions. The final theory is that of social ecology where the objective is to evolve an understanding of crime and social disorganisation through analysis of socio-geographical factors, investigating their role as explanatory variables (for example the relationship between crime and delinquency).

### Routine Activities

Cohen and Felson (1979) present their "Routine Activities" theory in which they argue that contact crime is a result of a convergence in time and space of three fundamental elements, namely; "*motivated offenders, suitable targets and absence of capable guardians against violation.*" (589). From this perspective the notion of locality (and time) is central in comprehending the mechanics and patterning of criminal activity. For example, research into the dynamics of commercial robberies (Van Koppen and Jansen 1999) indicated that the most robust explanator for their variations was that of suitable target availability coupled with their protective characteristics. It was concluded that this provided sufficient explanation, and a more complex justification was not required.

**Pattern theory**

Routine activities theory was developed further into pattern theory (Brantingham and Brantingham 1993) where interactions of people and places are modelled using the concept of nodes and pathways in space and time. People travel along known activity pathways to nodes around which the offender searches for potential targets. Similarly the potential victim's routine activities are of importance and also have the power to affect the patterning of crime. The theory identifies the importance of triggers, described as cues to the committal of a crime. For example, an accidental physical contact that leads to a spillage of drink in a pub may trigger an aggravated response (Brantingham and Brantingham 1981: 268).

Therefore, the theory posts that an understanding of crime patterning rests upon knowledge of activity and awareness spaces. Where there is a convergence in space and time of the three fundamental elements advocated by routine activities theory coupled with a triggering mechanism, then a crime is likely to follow. The theory permits an insight into the effect of these interrelated facets from which preventative measures can be deployed.

**Rational choice theory**

The rational criminal model (Cornish and Clarke 1986), with its roots in situational crime prevention (Clarke and Felson 1993), offers a complementary theoretical perspective to routine activities and pattern theory (Barr and Pease 1990; Gabor 1990). Rational choice theory was developed to offer an insight into the personal choices made by criminals. The emphasis of the theory is on the establishment of why the offender committed the crime. Committal of a crime is centred on opportunity versus perceived risks. For example, if an individual is walking along a road and sees an open car with a laptop on the front seat and the perceived benefits outweigh the risks then there is likely to be a theft.

The theory has been used to inform situation crime prevention programmes to reduce opportunity of crime by increasing the perception of risk. One example is that of a department store subject to a high level of shop lifting, with thieves typically entering the store grabbing an entire rail of clothing and running off. Risk reduction was achieved by the introduction of a simple countermeasure that was to decrease the portability of the clothing by reversing each hanger so they could not be removed in a single grab (Felson 2002).

Each of the three theories discussed above offer complementary perspectives from which the manifestation of crime can be conceptualised. These range from the broader societal picture (routine activities), towards local patterning (point pattern theory), ultimately the individual (rational choice) (Felson and Clarke 1998). Collectively and independently they offer an insight through which environmental design, situational prevention programmes and problem orientated policing can be evaluated, and new policy developed.

### **Social Ecology Theory**

In slight contrast to the previous place based theories, social ecology attempts to extract common social parameters to explain patterning of crime. The theory originates from the School of Sociologists based at the University of Chicago (Shaw *et al.* 1929; Shaw and McKay 1969). The work involved mapping the city of Chicago using various factors such as crime and social data across the various areas of the city. Through the aggregation of a variety of data (official and research derived), the patterns were sought. Similar work has been carried out in Britain (Burt 1925; Jones 1934; Castle and Gittus 1957; Wallis and Maliphat 1967) in which ecological generalisations of study environments were created and correlations established to types and levels of crime. More recently there has been a revival of interest in the original Chicago School ideologies (Anselin *et al.* 2000).

The social ecology approach has received some criticism, one example of which is that of the ecological fallacy (Robinson 1950). This is where generalisations were made on the basis of aggregate data that failed to represent the true underlying populations it was describing. Despite the criticisms received, social ecology has recently received renewed interest that can be partially attributed to the advances in GIS techniques (Anselin *et al.* 2000).

The advent of computing technology and the development of GIS technologies has undoubtedly fuelled the development of a range of crime mapping and analysis systems. According to Weisburn and McEwen (1995), the first crime mapping applications began to appear in the United States as early as 1967 (Pauly *et al.* 1967). Some of these early examples of computing mapping used the Synagraphic Mapping (SYMAP) package developed at the Harvard Laboratory for Computer Graphics and Spatial Analysis. SYMAP, although extremely primitive by today's standards, provided the first tangible evidence as to the potential of computers for efficient cartographic generation (a process that hindered the



cartographic process in the previous century). Current popularity of such technologies has seen the inception of the US Mapping and Analysis for Public Safety (MAPS) in 1997 for the development and dissemination of GIS-related research in the crime mapping arena.

### 2.3 Crime mapping and crime analysis

It is important to make a clear distinction between crime mapping and crime analysis. Crime mapping can be described as the mapping of spatially referenced data for visual inspection. Crime analysis takes the process beyond visualisation to involve further interrogation of the data set(s) involving statistical analysis of spatial phenomena, in which analytical procedures have been partially or fully automated.

#### 2.3.1 Crime mapping systems

Computerised crime mapping technology has been utilised to assist in the visual analysis and aggregation of data sources to create a snapshot of crime incidents. The use and development of crime mapping systems in the US is well documented (Mamalian and LaVigne 1999) with a quarterly publication detailing new developments and innovations (Police-Foundation 2002); use of crime mapping systems in the UK is less well documented (Ratcliffe 2000a).

Geocoding of police data has allowed the visualisation of crimes in a spatial context (prior to its analysis was restricted to text-based searches and pushpins in wall maps). The automation of this process has been discussed by (Openshaw *et al.* 1990) together with its usefulness, limitations and difficulties of implementation within an operational policing environment.

#### 2.3.2 Crime analysis systems

Development of crime analysis systems that have been explicitly designed to handle and visualise crime related data are a relatively recent phenomena. GIS were initially designed as tools for the storage, retrieval and display of geographic information (Fotheringham and Rogerson 1993). Their initial capabilities to conduct spatial analysis on geographic information were limited (Alber 1987; Goodchild 1987). Since these early days there have been many modifications that enhance the capabilities of GIS to conduct analytical operations on spatial data. The potential benefits for developing closer linkage between statistical spatial analysis and GIS have been noted by (Bailey 1994: 21) as:

- (i) *Flexible ability to geographically visualise both raw and derived data;*
- (ii) *Provision of flexible spatial functions for editing, transformation, aggregation and selection of both raw and derived data;*

- (iii) *Easy access to the spatial relationships between entities in the study area.*

Despite developments, standard off the shelf GIS products (such as ArcView, MapInfo and SERWorld) still offer relatively basic spatial analysis tools. In the context of crime analysis the tool sets tend to lack functionality for more advanced analysis such as clustering. This situation, however, continues to change with developments and extensions to the standard GIS tool sets including SCAS (USDOJ 1997) and the Crime Analysis Application Extension for ArcView (ESRI 1999), which incorporate bespoke tools to analyse crime related data within the GIS environment. To date, the majority of extensions have been written for ArcView, with much less software in general circulation for other systems. (For a fuller discussion of software availability for crime mapping and analysis see Police-Foundation (2000).)

In addition to extensions and customisation of standard GIS packages, there are stand-alone packages such as GAM (Geographical Analysis Machine) (Openshaw 1987), STAC (Spatial and Temporal Analysis of Crime) (Illinois-Criminal-Justice-Information-Authority 1988) and CrimeStat (Levine 2002). Each of these have been designed to readily interface with GIS packages and are capable of identifying spatial clustering. In addition, both STAC and GAM include temporal clustering techniques.

Use of geo-statistical tools in an operational environment is gaining momentum. In the US this is fuelled by the work of MAPS, which both promotes and provides funding toward their development. To date, UK police, fire and ambulance services have yet to realise fully the application of such technology. However, the uptake has begun and GIS has been applied to resource allocation in a command and control environment and basic mapping of incidents (see Chapter Three for a presentation of the results from a national survey of UK emergency services).

## **2.4 Geocomputational techniques**

Analysing the temporal geography of crime focuses on the identification of patterns. The identification of such patterning has brought about the development of numerous techniques. The following sections discuss a series of key geocomputation techniques:

### **2.4.1 Spatial distribution**

Techniques to analyse spatial distribution range in complexity, from simple descriptions of coverage (for example, the calculation of mean centre, standard deviation of the x and y co-

ordinates and minimum bounding area), to more complex computations (for example, the degree of spatial independence and degree of clustering). Here, the discussion has been divided into three sections -area, point and surface - each examining key techniques currently used within each domain.

### **Area**

A popular technique using area or aggregate data is the calculation of spatial autocorrelation. This test assesses the degree to which areas are correlated to one another. In the context of crime data a positive spatial autocorrelation refers to the situation where locales of high incidence adjoin other areas of similar high incidence. Conversely negative spatial autocorrelation is where incidence exhibit dispersion showing no evidence of clustering. Both the Moran I and Geary C indices (Moran 1950; Geary 1954) offer a measure for spatial autocorrelation that can prove useful in terms of assessing global trends.

A local form of Moran where the statistic is applied to individual zones has been developed by Anselin (1995). The local Moran is a local indicator of spatial association (LISA) (Anselin 1995) where the objective is to derive a localised estimate of connectivity and the degree to which it deviates from spatial randomness. The local form of Moran can provide a useful comparison to global estimates.

### **Point**

For the assessment of point data the Nearest Neighbour Analysis (NNA) (Clark and Evans 1954) is a commonly used technique. NNA measures the distances between all points and their nearest neighbour comparing it to a distribution expected by chance. The output is an index value that can be used to describe whether the point coverage tends towards clustering, random or a uniform distribution. The output value is typically supported with a significance statistic testing for first order nearest neighbour randomness. A limitation of NNA is the effect of edges on the calculation of the nearest neighbour. Thus, for points located around the periphery of the study region, nearest neighbour distance may be falsely exaggerated where closer points are located beyond the imposed boundary. The edge effect has however been considered as producing a conservative result where significant values are found on the basis that the majority of data exhibit some degree of clustering (Levine 2002).

### **Surface**

Derivation of a continuous surface from an underlying point-based distribution has become an increasingly popular technique to visualise concentrations of crime. Kernel density estimation is a popular technique by which these continuous surfaces can be generated. Placing a grid over the study area, density estimates are calculated for each cell using a three-dimensional function. Brunsdon *et al.* (2002) extends this technique using kernel weighting to generate localised summary statistics. These statistics then allow data distributions to be geographically visualised.

The Gi\* statistic (Getis and Ord 1992) - a LISA statistic, is capable of identifying local clustering based upon a global measure. The spatial non-stationarity nature of crime favours such local measures, with global measures having the potential to produce misleading results (Fotheringham *et al.* 2002).

The variety of techniques currently available, and the flexibility that many offer (for example, the options to determine class boundaries), allows the analyst to significantly influence the visual depiction of the data. Chainey and Reid (2002) attempt to consolidate some of these decision criteria by offering a formalised procedure by which statistically robust surface mapping can be consistently produced.

#### **2.4.2 Predictive techniques**

The geographical prediction of crime is a relatively new research domain that can be broadly broken down into two categories. The first is the prediction of individual behaviour, based upon an input pattern of associated events, known as geographic profiling. The second describes those techniques developed to forecast the risk of areas to volume crimes.

##### **Geographic profiling**

Geographic profiling (Brantingham and Brantingham 1981; Rossmo 2000) has been successful in both operational situations in the UK, US and Canada (ECRI-Environmental Criminology Research Inc 1999).

Here offender behaviour is modelled by measuring distances and routes travelled to commit crimes. The '*activity space*' (Brantingham and Brantingham 1981: 349) profile produced describes the routine activities of individuals. Using this regular journeying through geographic space, geographic profiling identifies the probable home locations of serial offenders (for example, repeat burglars and rapists). By

mapping the locales of crimes known to be associated with an individual and using a distance decay function, the offenders probable attack areas and residence are approximated. Results are then presented by a risk map surface from which search strategies have been founded.

### **Volume prediction**

Two theories are currently used; the first involves using the Broken Windows theory, the second incorporates the Rational Criminal theory.

The Broken Windows theory (Wilson and Kelling 1982) offers an insight into perception and crime fighting initiative stating:

*"...if a window in a building is broken and left unrepaired, all the rest of the windows will soon be broken...one unrepaired broken window is a signal that no-one cares...."(30)*

The theory stated that if smaller petty crimes were left unchecked, then this would fuel the contention that no one cares and would lead to an escalation in more serious crimes. Using neural networks and broken windows as a theoretical foundation, research into the formation and spatial prediction of emerging drug markets through identification of leading indicators has proved successful (Olligschlaeger 1997). Olligschlaeger (1997) applied a grid-based approach founded upon cellular automata to predict the emergence of new hotspots.

Both Broken Windows and the Rational Criminal theory inform crime and disorder prevention strategies through analysis of the properties of crimes (and offenders). Currently there are a limited number of MAPS funded research groups in the US working on this topic (Gorr and Olligschlaeger 1998; Rogerson 1998; Brown 1999).

## **2.5 Future trends and research focus**

The area of crime mapping and analysis remains both active and growing, with a variety of areas currently under investigation. By considering the motivational factors for carrying out future research in GIS-related crime studies, three types of research group can be identified.

- The first group will continue to develop new, and enhance existing, techniques for identification of patterning (for example, algorithms for the detection of hot spots and clusters, together with new techniques to handle spatially and temporally unspecific data

(Ratcliffe 2002)). In addition, there will be a closer union of GIS methods with other traditionally disjunct disciplines such as AI, data mining and knowledge discovery techniques. According to Brown (1998) use of AI techniques in crime analysis (for example, knowledge-based and expert systems) dates back to the late 1980's (Icove 1986; Coady 1987; Badiru *et al.* 1988; Ratledge and Jacoby 1989), although none of these incorporated spatial techniques. A more recent application (Brahan *et al.* 1998) involved work with the Hong Kong police department to develop an expert system. Unlike its predecessors this system incorporated both mapping and some spatial capabilities.

Data mining is a technique that for geographical research is a new tool that continues to grow in popularity. Data mining can be defined as a process of providing an automated extraction and presentation of patterns and knowledge from databases. Whilst GIS can be used to identify simple trends (for example the relationship between unemployment and violence), other not so obvious trends may lie undiscovered in the database. Data mining and knowledge discovery techniques have recently gained attention as a means of discovering such hidden information within databases. The GAM (Openshaw 1987) is an early example of a geographical data mining tool. However, to date the application of such techniques for the analysis of crime data has been limited (Brown 1998; ERA.Technology.Ltd 1999; ISL 1999).

The SMART Software Technologies Club (ERA.Technology.Ltd 1999; ISL 1999) have developed a system in conjunction with the West Midlands Police Force. The system utilises data mining techniques to identify similarities of crimes based upon physical descriptions and *modus operandi*, and incorporates a neural network trained to act as a "novelty detector" based upon the pattern and behaviour of an individual criminal. This technique evaluates the probability of a pattern being associated with a specific criminal.

- The second group is likely to be involved with the development and integration of new types of visualisation techniques and technologies such as global positioning systems (GPS), three-dimensional models, photography and satellite imagery. The integration of GPS technology potentially permits accurate identification of specific locations (for example, resources) in the field, in addition to allowing automatic geocoding of crime scenes in real time. These types of future uses are, in reality, going to suffer from well-noted limitations of GPS technology to function accurately and efficiently in urban environments.

The development of three-dimensional techniques has the potential to improve visualisation of the environment within which operations (for example, resource allocation decisions) and spatial analysis (for example, high resolution hot spot analysis) are conducted. The main issues concerning this type of research are likely to be related to model generation, accuracy, manipulation and efficient update of the model with new and amended features. In the context of crime studies research focus is likely to be concerned with the development of new tools for operational planning through spatial analysis.

- The final group will be involved with techniques and development of predictive models. A logical step beyond analysis of past and current crimes is the development of a predictive system capable of forecasting changes as well as the identification of fresh occurrences. Prediction in criminological research has been established for a number of decades (Ohlin and Duncan 1949; Mannheim and Wilkins 1955; Glueck 1960; Francis 1971). Foundation of predictive models with a geographic and GIS focus is still at a very early stage.

## 2.6 Summary

Crime is a social construct, and therefore some understanding of criminological theory (for example, routine activities and rational choice theory), is important in their interpretation.

A geographical approach to understanding the dynamics of crime in which GIS has played an important role has become increasingly popular. A significant amount of research has been conducted on the implementation of geocomputational techniques for the interrogation of archive crime data to reveal patterning. However, few methods currently exist for the prediction of future patterns.

### **3. COMPUTERISED MAPPING AND SPATIAL ANALYSIS AS A STRATEGIC AND OPERATIONAL TOOL**

*This chapter discusses the results of three national surveys sent to each of the UK's Emergency Services (ES) to quantify the use and uptake of computerised mapping technologies. The aim of this research was to highlight current trends, attitudes and implications of such technologies in addition to the identification of potential future developments. Results from the surveys provided a grounding from which additional techniques were later developed.*

---

#### **3.1 Introduction**

Geography plays a vital role in a multitude of operational and strategic procedures for each of the three ES. These include operational issues such as resource location and deployment via the most efficient route. In addition, strategic issues, such as demand evaluation, intimately involve a variety of geographical parameters that can be informed through the utilisation of computer mapping and GIS. As a result, the correct application of computerised mapping and related technologies could potentially enhance the overall performance of ES. In many cases this may result directly in enhanced reaction capabilities in time critical services.

To date there is limited information as to the degree to which computerised mapping has been adopted, integrated and utilised by UK ES. Openshaw *et al.* (1990) documented how early attempts to establish an operational GIS within one particular police force resulted in failure. Reasons given by the police for this included system reliability, operability, user friendliness and vendor support issues (pg. 2.3.2). In terms of a national perspective, only the police force had been evaluated (Ratcliffe 2000a), however that study was founded upon a telephone interview, with the published results mainly providing insight into attitudes, thus offering limited technical information. The survey did, however, expose certain technical issues concerning the appreciation and implementation of user requirements, consistent with problems raised by Openshaw some ten years earlier. The aim of the surveys presented here is to provide a detailed quantification and qualification of its application.



Results from the surveys offer an understanding of the current technologies, issues and attitudes that in turn provide a grounding from which additional techniques (detailed in subsequent chapters) are later developed.

### **3.2 Mapping Systems in the ES**

The development of GIS and spatial technologies, have undoubtedly fuelled the proliferation of a range of mapping and analysis systems. The flexibility, powerful visualisation and integral capacities that these systems offer have greatly assisted in their utilisation. Deployment of mapping technologies within the ES can be divided into two broad categories, namely time and non-time critical systems.

- Time critical (real-time) applications incorporate additional technologies such as Global Positioning Systems (GPS), Automated Vehicle Location (AVL) and mobile data devices. Routing algorithms typically support such systems, where they are used to manage and monitor vehicle allocation and dispatch. Advanced routing techniques are able to account for traffic flows and road blockages when determining the most efficient route to an incident.
- Non-time critical applications include a variety of visualisation, interrogation and analysis techniques using archive data to provide an assortment of spatial and statistical output that can be utilised to evaluate, direct and assist the creation of future policy. Output from these non-time critical applications can be utilised to assist in the future direction of real-time operations (for example, resource allocation decisions based upon an interrogation of archive data).

The following sections discuss each ES, providing a historical perspective in terms of their involvement with the mapping process in relationship to their operations.

#### **3.2.1 Police**

The UK policing operation is geographically segregated into ten regional areas consisting of 52 Constabularies, each of which is further sub-divided into a series of territorial divisions under Constabulary jurisdiction. On average, each Constabulary consists of six territorial units employing approximately 3,250 police and civilian staff policing a population of approximately 922,000.

Recently, the use of paper-based mapping techniques have been replaced and extended through the use of computer based crime mapping technologies. Their use are rapidly becoming a vital prerequisite to understanding incident distributions, assisting in the identification and allocation of resources, and the development of policing strategies. The ability to efficiently generate simple maps to depict crime location and densities can be used directly to inform policing strategies, therefore increasing officer effectiveness. A recent report published by the Home Office (2000a) has underlined the importance of geographic data for the analysis of crime at the local level. It states that through its use, police forces will be able to conduct a more detailed analysis of their data to identify hot-spots and thus improve the allocation of resources. In addition mapping is advocated as a tool to identify incidents occurring across the various police borders.

### **3.2.2 Fire**

Ten fire service regions cover the UK, within which there are 58 Brigades, which on average consist of three divisions (territorial areas). Each employs an average of approximately 1,300 fire and civilian officers, typically serving a population of approximately 1,160,000.

The emergence of computing technology in the UK fire service began in the early 1990s with the introduction of desktop computing, and more latterly GIS employed for planning resource location and distribution - known as the Fire Cover Model (Reynolds 1998). The model facilitates the utilisation of locational (station and incident localities) and network (road) data for fire cover planning in order to optimise resource usage. In addition, information on occupancy, building layouts, hydrant and sprinkler systems, and the presence of potentially toxic materials can be combined to provide a clearer indication of incident parameters. Such information can subsequently be relayed readily to responding vehicles via onboard mapping terminals.

In addition to responding to fire related incidents the fire service is also required to attend *special service calls* that include natural disasters and road accidents. Similarly, these incidents demand comparable geographical variables to be assessed in order to optimise efficiency, thus GIS potentially have an important role to play.

Literature detailing the implementation of dedicated fire service GIS is extremely limited. However, the majority of published reports are in electronic format, sourced through individual Brigade web sites.

### **3.2.3    Ambulance**

The UK ambulance service is geographically segregated into ten regional areas consisting of 34 Services across England, Wales and Northern Ireland, in addition to six Scottish Service regions. On average, each service is composed of 26 stations employing approximately 925 staff, serving a population of approximately 1,700,000. For convenience the term service is used hereon to collectively refer to both trusts and services.

In common with police and fire deployment, the ability of an ambulance team to respond to an emergency call is a function of a series of spatial, temporal and contextual factors, each contributing to its efficiency. The more efficiently an ambulance team can respond to an incident the better, thus the challenge is one of strategic resource allocation and vehicle routing. In addition to emergency responses, ambulance services are responsible for less time critical operations, such as patient transport to and from hospitals. Such movements also demand consideration of spatial, temporal and contextual factors in order to maximise efficiency.

Reports of GIS-based applications within the ambulance service began to emerge in the early 1990s in the US (Barry 1991; GIS-Newslink 1993) and a little later in the UK (Cattini 1997; Smith 1997). However, there is relatively little material published on the subject. The development of frameworks and methodologies for determining ambulance location and deployment has been the subject of some research (Peters and Hall 1999; Derekenaris *et al.* 2000). Beyond time-critical applications, however, reports of applications and techniques for interrogation of archive data are extremely scarce (Carcach 2000).

### **3.3    National surveys**

The three surveys undertaken each targeted a level that would best provide a comprehensive overview of all computerised mapping activities at both local and regional levels of jurisdiction, and in the context of national policy. This equated to the Constabulary, Brigade, and Service level for the Police, Fire and Ambulance respectively. The survey tool (a modified version of the questionnaire used in a US crime mapping study) was designed to pose regional-wide questions in relation to generic activities, strategies and future policy in addition to previous, current and perceived future engagement with computerised mapping. The survey tools used for each of the ES are included in appendix A.

Overall the survey generated a **52%** (Police), **54%** (Fire) and **54%** (Ambulance) response rate equating to **27** Constabularies, **31** Brigades and **19** Services respectively. A total UK

coverage was not achieved, with an indication from some non-respondents that either a lack of time or a reluctance to disclose certain intelligence information.

Due to the proportion of missing data, the results presented are used as a guideline to determine the national picture. One of the main questions asked was whether or not *“the respondents are reflecting an unbiased view of the national picture, or indicating an unrealistic majority of pro-GISers?”* Existing literature, although scarce, coupled with the comments of respondents offered some additional information. This was developed further through several personal communications, with both respondents and those not wishing to participate in the survey helping to account for a portion of the incomplete data.

### **3.3.1 Mapping or No Mapping**

Acquisition of computerised mapping software was shown to be a relatively recent occurrence for all ES, with a rapid increase in use within the past five years (Figure 3.1).

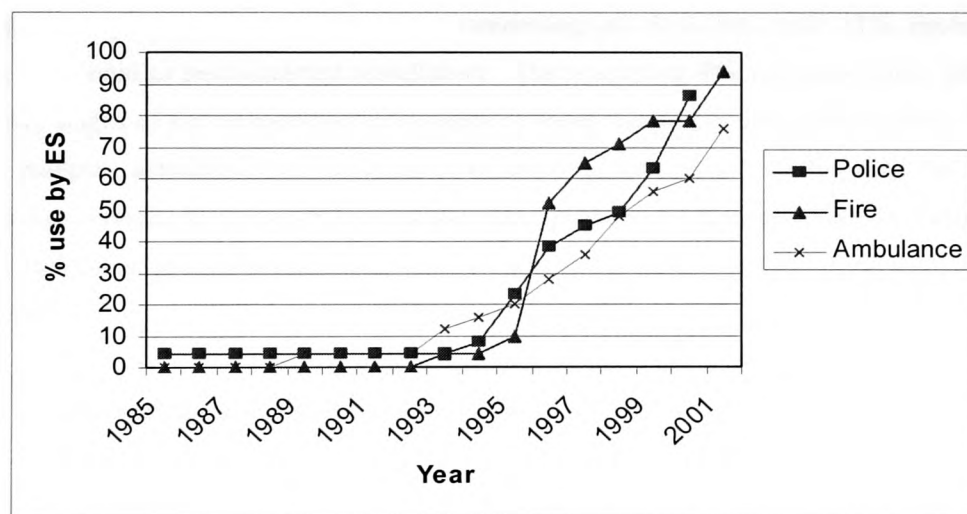


Figure 3.1 Uptake of mapping technologies timeline

In relation to the ES respondents, the large majority (85% Police, 94% Fire, 76% Ambulance) are currently engaged in some form of computerised mapping. A 1998 US survey of crime mapping by police agencies indicated a figure of 13%<sup>1</sup> of departments engaged in computer based crime mapping, contrasting to a UK figure of 48% indicated by this survey at the same point in time (1998 figure). The disparity between the US and the UK

---

<sup>1</sup> Statistics taken from Mamalian, C. A. and N. G. LaVigne (1999):1

usage represents the department level at which the US survey was targeted, where the majority represented agencies employing less than 100 officers. A figure of 36% was reported for those departments employing in excess of 100 officers, providing a more comparable sample to that represented in the UK survey. Unfortunately, no international comparisons for either ambulance or fire services currently exist.

For the police, of the **85%** currently using digital mapping, a total of **44%** acquired their computer based mapping technology since 1998 and **91%** since 1994. Currency of acquisition was less pronounced in both the fire (**23%** since 1998 and **90%** since 1994) and ambulance service (**28%** since 1998 and **60%** since 1994). For those not currently engaged in computerised mapping, all intend to invest in such technology, **50%** of which plan to do so within the next twelve months, the remainder within a 1-2 year period (ambulance services did not specify).

For those Forces engaged in computerised crime mapping activities, **74%** have implemented a Constabulary wide mapping facility incorporating all divisions, with **22%** deploying a single divisional or head-quarters installation. The remaining **4%** represent either Forces in the early stages of the technology deployment or those wishing to centralise all computerised crime mapping activities. The deployment of mapping technologies within both the fire and ambulance services are largely located within headquarters installations (fire **90%**, ambulance **95%**) reflecting the particular focus on centralised resource identification and allocation activities.

### **3.3.2 Information output & dissemination**

Successful integration of any mapping software within an operational environment is reliant upon both technical and strategic knowledge, coupled with a comprehensive plan for its exploitation.

The survey indicated that the police were the largest employers of specialised staff to conduct computerised mapping activities (**91%** police, fire **59%** and ambulance **53%**), **43%** employ both specialised staff and police officers (fire **31%** and ambulance **5%**) and **9%** of Forces utilise only police officers. The ambulance service were the largest users of their own staff, utilising **47%** to perform their mapping (fire **17%**).

Key to the effective use of computerised mapping is the dissemination of its outputs that can take the form of a point, area or hot-spot maps to resource managers and serving officers.

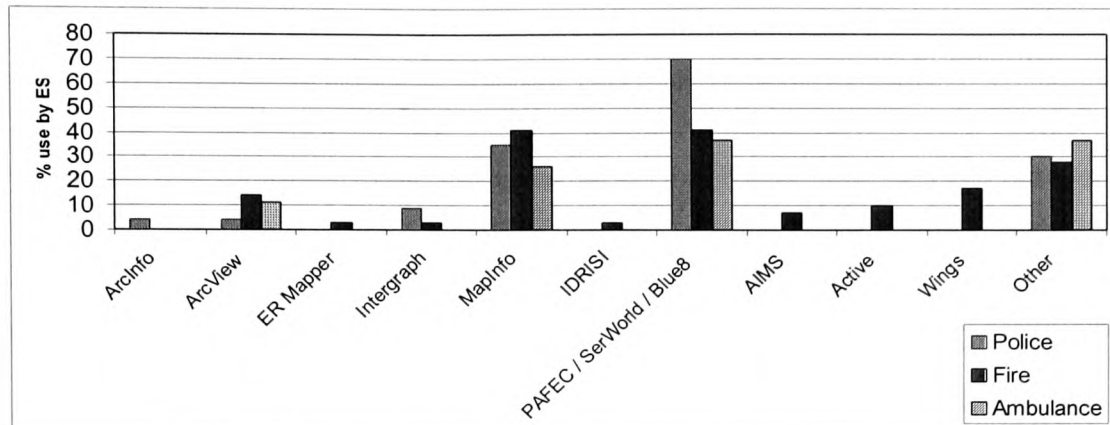
One potential mechanism for achieving this is through the use of a secure Intranet that offers an efficient, low maintenance and economical method for dissemination to all necessary parties. The survey indicated that all Forces possess an Intranet (43% of Forces reporting that in excess of 70% of their staff have direct access to such facilities) although only 35% utilise this facility to disseminate output from their mapping activities. The existence and use of intranet technologies in both the fire and ambulance services was found to be more limited (fire 66% and ambulance 53%) of which only 17% (fire) and 5% (ambulance) used it for mapping dissemination.

The regularity with which the analysis of command data is conducted can potentially impact upon its effectiveness as a tactical tool and its ability to convey accurate, timely intelligence concerning the status of a region. This was reflected in the survey whereby 78% of police, 58% of fire and 58% of ambulance services conducted analysis daily. 13% of police, 3% of fire and 11% of ambulance services carried out weekly analysis, and 9% of police, 28% of fire and 26% of ambulance services on a monthly basis. The currency of the spatial data and cartographic output is of particular importance for tasks such as informing officers (of which 87% police, 69% fire and 53% ambulance do so) and to a greater degree for identification of resource requirements such as command and control (of which 65% police, 52% fire and 100% ambulance are involved), where response within relatively short timescales is essential. The demand for real, or near real-time, mapping information is reduced somewhat when applied to other areas of ES operations (such as evaluation of policies and strategies, as carried out by 52% of police, 48% fire and 58% ambulance, reflecting, to some extent, the numbers of ES generating less regular monthly mapping output and spatial analyses).

### **3.3.3 Software selection**

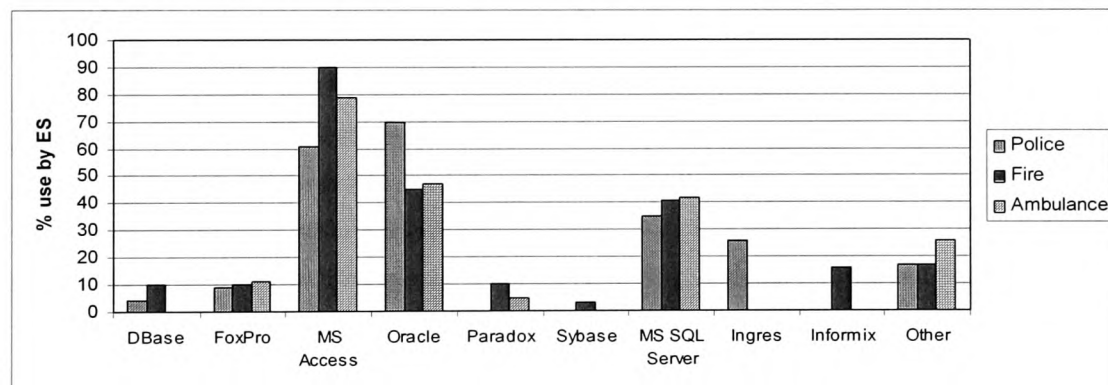
Selection of mapping software was shown to be somewhat divided with 48% of Forces, 48% of Brigades and 26% of Services utilising two or more packages; with SER products (PAFEC, SERworld, Blue8) and MapInfo® proving the most popular, while a relatively large number opting for a miscellaneous selection of packages (Figure 3.2). For Forces utilising more than a single software provider, no specific combination proved more popular.

A major cost consideration in the selection of mapping software is its ability to interface with existing data sources, demanding both minimal modification to hardware/software and alteration to existing working practices. One major consideration is the database (or in many cases databases) storing the information to be mapped, and their compatibility with the mapping software. The survey revealed that the most popular database was Microsoft Access



**Figure 3.2 Mapping software selection**

(61%-Police, 90%-Fire, 79%-Ambulance) followed by Oracle (70%-Police, 45%-Fire, 47%-Ambulance) and Microsoft SQL Server (35%-Police, 41%-Fire, 42%-Ambulance). The majority of ES (70%-Police, 76%-Fire, 58%-Ambulance) possess multiple types of database software, with a common configuration being a corporate Oracle database and instances of Microsoft Access for local access and data manipulation (Figure 3.3).

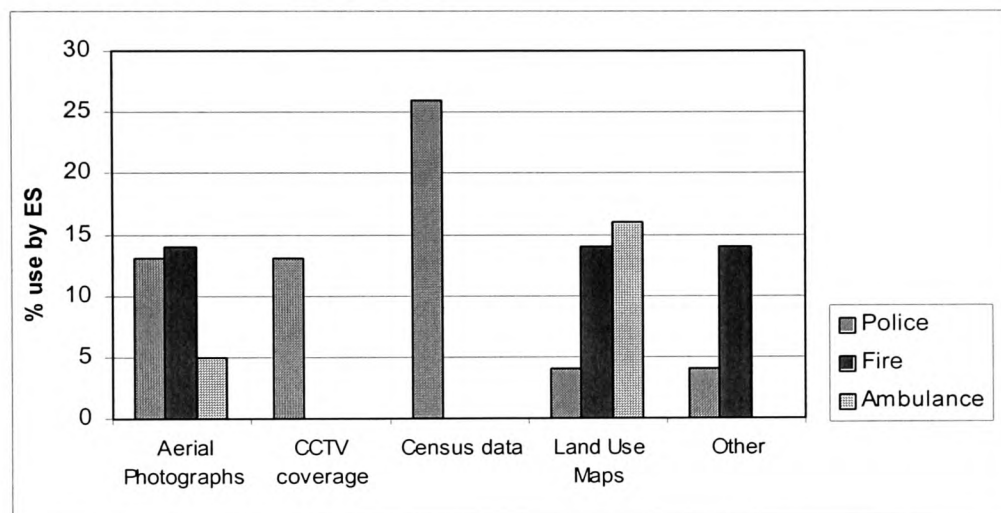


**Figure 3.3 Database software selection**

A fundamental function of any GIS is its ability to offer the integration of disparate data sets, enabling visualisation to provide a broader picture. The value of this functionality lies in its potential to offer a valuable insight into the patterning of crime through aggregation of various data sets. For example, the ability to visualise an incident locality in relation to land use and physical structures is potentially a valuable tool. It was shown that all ES utilise some form of street map data, including an assortment of Ordnance Survey (OS) products (1:1,250, 1:2,500 and 1:10,000 raster tiles, AddressPoint® and OSCAR® road centreline

data), Automobile Association (AA) map data and a minority (4%-Police, 7%-Fire, 10%-Ambulance) involved with in-house development of their own base map data.

In addition to street map data, other contextual data sets such as aerial photographs and census related information can provide a valuable insight for the analyst. However, the survey showed limited use of such information (Figure 3.4); it is thought likely that the cost of acquisition, integration and maintenance of these various data sets currently outweighs their perceived benefits for the majority of ES.



**Figure 3.4 Contextual data usage**

### **3.3.4 Mapping and Analysis techniques**

A fundamental dichotomy between the various mapping systems throughout the ES are those involved with purely mapping and those that additionally include analytical capabilities.

Mapping activities can essentially be divided into four categories, where **all** Forces and **90%** of Brigades map offence data. Calls for service data is mapped by **74%** of Forces, **69%** of Brigades and by **all** Services. Road traffic accident data is mapped by **30%** (**79%** fire) map road traffic accident data and **17%** involved in miscellaneous mapping endeavours. The survey identified four key techniques for the analysis of such data:

- **Automatic pin maps**

The ability to generate simple point-based mapping depicting the individual locations of resources and incidents is a powerful visual tool, which potentially can be readily updated and refreshed with new information. Output (dependent upon the functionality of the mapping software) can be efficiently tailored according to purpose; for example, mapping



can be produced to represent a specific type or types of incident, incidents occurring within a specified distance of a certain type of building, or for a specific snapshot in time. Automatic pin-maps, therefore, offer a highly reusable format by which core information can be queried, aggregated and presented in a variety of styles according to purpose. Its popularity as an ES tool is reflected by 57% of Forces, 17% of Brigades and 47% of Services utilising this technique.

- Hot-spot mapping

Hot-spot mapping, when compared to the automatic-pin map technique, involves a greater degree of data processing, but arguably results in the production of a cleaner portrayal of incident volume. The aggregation of point data into a single colour-coded density surface, distinguishing areas of high, medium and low crime levels, provides a simple visual depiction of a region's vulnerable areas. Its popularity was clearly demonstrated by the survey, with 83% of Forces, 69% of Brigades and 79% of Services utilising hot-spotting techniques.

- Temporal analysis

Temporal analysis also represents a relatively popular technique, with 48% of Forces, 24% of Brigades and 42% of Services conducting some form of temporal interrogation of their data. This may constitute the ability to isolate specific snapshots in time, such as a specific day, date and time. Spatial output can then be created to represent the incidents at that time, and can provide a valuable insight into the temporal effects upon resource demand.

- Other Techniques

A range of other techniques were reported in the surveys, which accounted for 17% of Forces, including predictive and profiling procedures. Although both procedures are currently less well exploited, it is likely that as techniques become more established and embedded within existing practices they will receive wider acclaim. 14% of Brigades reported use of alternative techniques including *add hoc* queries driven by strategic demand. 11% of Services reported the use of time mapping techniques that, presumably, are used in a predictive manner to optimally locate resources.

### **3.3.5 Customisation and Global Positioning Systems (GPS)**

Effective deployment of mapping software has in certain cases necessitated customisation of the software. The survey suggested that 74% of Forces, 62% of Brigades and 56% of

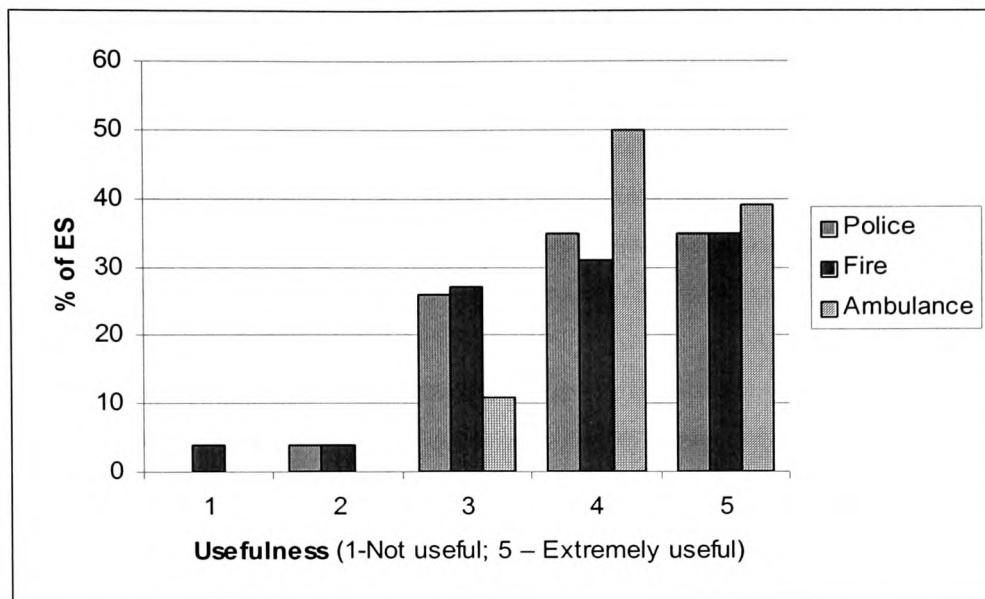
Services, found it necessary to customise their standard mapping software in order to integrate them with existing databases (39%-of Forces, 38%-of Brigades, 58%-of Services), provide specialised visualisation (4%-of Forces, 28%-of Brigades, 58%-of Services), or querying tools (43%-of Forces, 38%-of Brigades, 42%-of Services). Of those that customised their software, some (76%-of Forces, 61%-of Brigades, 50%-of Services) resourced the development in-house, while the remainder presumably utilised external contractors to complete the development task.

Location in the form of a spatial reference is key to mapping applications. The ability to identify a specific locality as a resource or incident is of prime importance. The integration of GPS technology offers powerful visual and analytical capabilities through the provision of a spatial reference in the form of an x,y co-ordinate. Accuracy is dependant upon equipment, pre and post-processing techniques and environmental characteristics. However, the technology can prove useful for identification of resources and localities of incidents.

The use of GPS technology within the ES environment is not new, and a range of systems have been implemented. This survey, however, highlighted that its use in an operational context is not widespread by both police and fire (30%, and 24% respectively). Ambulance services were the greatest user, where 68% currently use such technology for a combination of resource identification and allocation activities. It is likely that this figure is set to increase in future years, especially with the fall in cost of GPS technology, coupled with the decision to remove Selective Availability. As of the 1<sup>st</sup> May 2000, typical commercial GPS position accuracy is approximately 20 metres (opposed to 100 meters prior to this date).

### **3.3.6 Perceived usefulness of computerised mapping**

Perceptions of the usefulness of mapping technologies were generally positive, on a scale from one (not useful) to five (extremely useful), 70% rated either four or five (police), 66% (fire) and 89% (ambulance) (Figure 3.5). These figures were backed by further comments stating that *"it is clear that this service will become a vital tool in the case of fire service activities"* and *"there is a great deal of interest in crime & disorder mapping within the Force."* Despite the positive majority, some concerns noted software functionality and usability as detrimental factors to its usefulness in both operational and strategic arenas.



**Figure 3.5 Perceived usefulness of computerised mapping**

### **3.3.7 Implications, problems and looking to the future**

Successful application of mapping technologies into pre-existing operational environments raised a number of concerns. The survey attempted to identify a series of key problems, and the degree to which they impact upon effective application of computerised mapping. A series of topics (as identified in the US survey) were evaluated according to the extent to which they created a negative impact upon the ES and its mapping activities using a scale of one (No problem), to five (Serious problem), under the following headings:

- Limited computer resources
- Limited financial resources
- Limited time
- Limited training opportunities
- Limited working knowledge of how mapping is used in the field
- Limited interest from administration
- Limited interest from support staff
- Difficulties with computer software

In general the majority of replies across each of the ES were similar, identifying common weaknesses and strengths. Certain questions, however, resulted in a spread of opinion within each ES, that to a certain extent reflects the variety of approaches each has taken. This was

particularly evident in response to limitations of working knowledge of how mapping is used in the field by fire brigades (52% rating one or two and 35% rating four or five). The continuity of staffing may to some extent explain this division of opinion, with some brigades subject to higher throughputs of personnel, for example (*“transfer of post bring with it a loss of valuable intimate knowledge experience of the systems resulting in a high repetition of questions asked”*). Police forces and ambulance services also indicated concerns on this topic (48%-of Forces, 37%-of Services rating four or five), with personnel relocation one possible explanation for this.

Limited computer resources, interest from support staff, and difficulties with computer software were least problematic. Both financial resources and time constraints were of greater concern. The limitations in training opportunities were rated highly problematic by both the police and ambulance (35% and 43% respectively), scoring this as either a four or five, with fire brigades offering a more neutral picture (69% rating two or three).

Additional problems reported concerned the validity of the spatial data that influenced the system accuracy, and ultimately user confidence, for example (*“Our data is appalling. Hydrant location is critical to operational planning & activity. We have touch-screen PC’s on our incident command vehicles, which show erroneous information. As a result there is little confidence in the tool”*). It is likely that this problem is more widespread than that highlighted by this survey. The remedy to such a problem lies in the instigation of stringent data collection, manipulation and meta-data standards in an attempt to achieve “cleaner”, more accurate data.

Central to the success of a computerised mapping installation is its integration to fall in line with corporate objectives. In some instances the survey revealed that some installations lacked this direction, for example (*“We have no overall strategy regarding GIS – it’s a fun toy!”* and *“they see GIS as producing ‘pretty maps’”*). As a result this hinders the realised and perceived benefits, and may as a consequence affect future support. However, given recent motivations and Government policy for ES (in addition to other Agencies and Organisations) to work in partnership this is likely to change. This will, no doubt help to focus attention on their better exploitation as their potential to help tackle broader community issues are realised (see Chapter Four). This, in turn raises another potential problem highlighted by the surveys, regarding the multitude of mapping configurations, and the danger of non-connectivity between components within and external to the parental organisations. This, to some extent, is reflected by the percentage of those currently engaged in cross-constabulary, brigade and

service collaborations (13%, 7% and 5% respectively), where technical (in addition to political) barriers preclude joint initiatives. This will become a greater issue, especially when considering the drive to work together more closely in Partnerships and the establishment of data sharing protocols. Herein lies the potential for further research.

In relation to crime mapping, the recent establishment of a list server (CRIMEMAP - MAPS 2003), dedicated to the discussion of crime mapping issues has certainly started to challenge this connectivity issue. The main thrust of the list server is to resolve specific problems by discussion amongst the registered users, through the application of common techniques, shared knowledge and experience. At the time of writing, US police crime analysts and university-based researchers mainly use the service; a small number of UK based practitioners also partake in discussions. The common threads of discussion range from questions concerning suitable file formats and conversion types to those involving a more lengthy examination of the most suitable techniques for specific problem domains. For example, deciding the most applicable techniques to identify evidence of displacement following the instigation of a targeted initiative. The service offers a valuable knowledge resource for the cross fertilisation of ideas to a growing number of individuals. The format of CRIMEMAP would also be applicable to both the fire and ambulance services, although no comparative facility currently exists. Participation in such a service obviously demands access to an Internet connection, which for some ES would require modifications to their existing IT infrastructure.

### **3.4 Summary and Conclusions**

The current indication is that Governmental support for computerised mapping technologies is set to continue. The importance of geographic data for analysis of police data is defined in a recent Home Office report (2000a) and by an earlier telephone survey (Ratcliffe 2000a). Survey evidence presented here support this attitude. The Survey also suggests that similar sentiments exist within the fire brigade. This is backed by a range of risk assessment toolkits produced by the Home Office (Wright 1997; 1999a; 1999b) that position GIS as an integral component in operational exercises. Similarly, the results indicate that this is also true of the ambulance service, however, no publications exist on this matter. For those ES not currently engaged with computerised mapping, the Survey suggested that all included would do so within a two-year period.

In summary, this Survey has shown both a growing influence and positive sentiment toward the use of computer based mapping within the UK's ES. The overall perception of

computerised mapping, despite various concerns, is that it offers a valuable tool to both operational and strategic functions. Financial support from leadership in terms of resources is set to continue.

Results from the surveys suggested that each ES exhibited several areas of commonality (for example, software selection and use of specific techniques). However, there remains a large diversity of approaches, which to a certain extent is shown by the degree to which each has found it necessary to customise standard configurations. Therefore, for greatest application in an ES computerised mapping arena, additional techniques that are developed should aspire to be both transparent and flexible (for example, platform and software independent). For some ES, these techniques should attempt to address the apathy that currently exists through the development of targeted tool sets. These tool sets could then provide a direct input towards decision making (for example, reducing the burden of producing cartographic output through improved data cleansing and geocoding procedures).

## 4 ETHICAL DILEMMAS IN CRIME AND DISORDER ANALYSIS: AN OVERVIEW

*This chapter presents a rigorous ethical evaluation that the study underwent at its time of inception. Details of the research proposal are presented, followed by the series of ethical implications that arose subsequent to applications being made to guardian organisations for access to their data archives. The chapter concludes with suggestion of a partial appeasement to the ethical dilemmas.*

---

### 4.1 Introduction

The role of ethics in the use of crime data is complex. The research presented in this chapter draws upon this thesis as a case study to exemplify some of the dilemmas that are typically faced by both the ethics committee and the researcher. Resolving such dilemmas is not simple. However, certain opportunities exist through which barriers can be overcome. The chapter concludes with the suggestion of several fronts by which a partial appeasement could be achieved.

### 4.2 The ethics of crime analysis and prediction

One of the most powerful facilities of computer based modelling is the ability to cross reference and compare different, previously disjunct, data sets (Ekblom 1988). Bowers and Hirschfield (1999) underline this point, stating that utilising GIS in the context of crime studies:

*“enable links to be established and spatial relations to be explored between data derived from different sources (e.g. crime reports, census variables, transport information and land use) and where appropriate data are available (e.g. grid-referenced individual level crime records) analysis can be undertaken to overcome the confounding effects inherent in the use of spatially aggregated data for pre-determined geographical boundaries (e.g. police beats).” (161)*

The ability to link and data mine vast, disjunct, data sets offer immense potential, the potential value of the combined data often being of far greater worth than the summed value of the separate entities. However, the anxiety of many is that these data sets can be combined to form an electronic alter ego, whereby individuals and areas can be pinpointed and unfairly translated. Such concerns extend beyond the confines of crime data.

*"A standard response to concern about increasing surveillance by the state is that 'if you have done nothing wrong you have nothing to fear'. A little over sixty years ago the Netherlands had a stable, open regime that collected extensive data on its population to enable efficient provision of services. It seemed that if someone had done nothing wrong, they had nothing to fear. Upon invasion from Germany by the Nazi regime, that data rather than being used to enable efficient provision of services, was used to enable a more efficient program."* (Fairwater 2000: 1)

This perceived dualism can be exemplified through numerous situations where on the one hand a data set can be utilised for the good of the community or individual, and equally, in contrast, can be detrimental when utilised in a certain manner. Thus, the use of low level socio-economic information for a given region can be utilised beneficially by a resource planner, such as the police or local authority, but can also be utilised as a means to negatively label that same area. Labelling has traditionally been applied to individuals, but there has also been suggestion of spatial negative labelling (Ratcliffe, 2000b). Therefore to develop the previous example, labelling certain areas potentially promotes a situation where the inward flow of positive influence (for example, investment in housing, amenities and businesses) becomes curtailed. Simplistically, the removal of positive influences may then act as a vacuum filled by an inward flow of negative influences (for example, poorer public transport facilities and the deterioration of the urban fabric). In essence, spatial negative labelling can be compared with evidence promoted by the Broken Windows Theory (Wilson and Kelling 1982), whereby a downward spiral effect might be induced in an area labelled as high in crime.

For example, in the United Kingdom, the labelling of certain schools as "failing" has often exacerbated their problems. Existing teachers move to better schools, leaving unfilled vacancies, and concerned parents send their children to better schools. The labelling can become a self-fulfilling prophecy - very quickly the label "failing" can be replaced with "failed."



A local example from the study region was a proposed scheme known as “naming and shaming” that planned to identify public houses and nightclubs where violence regularly occurred. Combining data from the Accident and Emergency departments of local hospitals and data concerning violent acts from the Police, the scheme planned to publish the results through the local media. Despite the hailed success of a similar scheme in Australia, the implementation of such measures, where the benefits are obvious, may, however, induce negative influences - ultimately becoming as with the schools example, a self fulfilling prophecy. On the one hand both officials and the media saw the desire as operating a safe, violence free establishment, which promotes proactive policy. On the other hand, however, local licensees may feel unduly compromised by such pressure. In addition, such a scheme is fallible to incorrectly targeting establishments as the locale for violent events, the control and avoidance of which may be beyond their influence. As with the schools, this could instigate a downward spiral, whereby existing public houses and nightclubs have their licences revoked, instigating a reduction in the prosperity of such targeted regions.

Similarly, the identification of high crime areas poses obvious problems in terms of labelling. However, its use in terms of promoting positive influences should be the foundation of measurement. Thus, the ethics of the labelling process should evaluate and control the way in which the labels are derived and used as a prerequisite to any form of analysis or dissemination.

### 4.3 A case study

The following is the original PhD proposal for this research. It has clearly stated objectives that, if realised, would have tangible benefits for those that might otherwise be the victim of crime. Nonetheless, it failed to gain ethical approval. An example of the ethical conundrum.

The proposal had the creditable aims of carrying out a spatial analysis of inter-personal violence. It was envisaged that the work would enable four key questions to be addressed: Where does crime take place? When does crime take place? Why does the crime take place? What can be done to reduce the crime rate?

*Inter-personal violence has increased markedly over recent years, and is now the most common cause of fractures of the facial skeleton in Western Europe. Crime rates differ between different types of urban district, and these differences are best explained by the variation in use of urban sites by differing populations. Whilst a database of violent incidents (location, cause, victim details including injuries and address) is rich in spatial*

*information, studies to date have been limited to simple statistical analyses of these variables. However, a much richer survey can be undertaken by linking this database with other databases, such as the Census of Population, police records, weather patterns, and accident & emergency data. Such information can tell us a lot more about the environment in which the incidents take place; and the home environment of the victim and/or perpetrator. Whilst geographical information systems (GIS) can be used to identify simple trends, for example the relationship between unemployment and violence at different levels of granularity, other not so obvious trends may lie undiscovered in the database. Data mining techniques have recently gained attention as a means of discovering such hidden information within databases. One of the more popular approaches to data mining utilises neural networks to detect trends.*

*The project will entail implementing a prototype GIS which integrates the various spatial databases discussed above to analyse and map the identified trends. The analytical engine will be supplemented with a neural network for some of the processing.*

The project received unanimous ground level support; the potential benefits were clear, and so requests to ethics committees were progressed. Unfortunately the proposal was refused ethical approval. The following section details the concerns that were raised.

#### **4.4 Conflicts between case study proposal and ethical approval**

One of the problems with the proposed project is that it suggested the use of numerous data sets – collected for reasons not related to the proposed analysis and, often, for use only by the guardian institution – that fall under the auspices of different ethics committees. The question these different ethics committees are charged with answering is not “is this proposal ethically right?” or “is this project of benefit to society?” but “does the proposed use of the data (we are charged to protect) conform with the purpose for which it was collected?” Where the answer to the latter question is no, then, however commendable the proposal, options open to the committees are limited. Drawing on ethical policy employed by the University of Glamorgan, the following discussion raises a series of key topics highlighting the reasons for the proposal’s failure.

Firstly, the proposed usage of the data bore no relation to the reasons for which it was initially collected, thus informed consent for use in an analysis exercise was beyond the remit of the historical data. Any attempt to contact each data subject was deemed impractical, in addition to the likelihood that some would probably not be willing to be included. Another issue is

that of confidentiality of the data subjects, with the proposal detailing the necessity of a spatial reference from which vulnerable areas could be identified. The use of this spatial reference was proposed to be in the form of a postcode, so not to identify any individual. The research did, however, propose the use of additional data sources, which in conjunction with a postcode may reduce the anonymity of data subjects. A final issue is that related to the avoidance of harm, with the proposal aiming to identify vulnerable areas with suitable intervention schemes to follow. However, the issue of negative labelling, discussed earlier, presented unresolvable concerns.

#### **4.5 Appeasement of the conundrum**

Following consideration of the various ethical implications, a tentative appeasement of the conundrum can take place on at least four fronts: Front 1, a reduction in the demands made by the researcher; Front 2, the liberation of ethics committees from some of their constitutionally imposed rigidity; Front 3, the introduction of a general principle that labelling will only be used to ameliorate what is problematic in an area, and Front 4, the use of technology to facilitate anonymous “black box” joining and interrogation of data.

- **FRONT 1: Reduction of data content demanded by the research.**

The proposal aimed to achieve the maximum benefits feasible given current technological and methodological statuses. However, a major problem encountered during the ethical evaluation was the use of a spatial reference (i.e. a postcode) potentially detailing localities of individuals, especially when used in conjunction with additional contextual sources – where issues of data protection and human rights are of concern. Appeasement of this could be through the agreement of spatial and contextual information standards, whereby data can be ethically disseminated, and issues of data privacy not compromised. This could take the form of a trimmed spatial reference (for example, using a sector level as opposed to a unit level postcode) whilst still maintaining all the descriptor information. Although the identification of risk and vulnerability is compromised to some extent, the overall aims of the research are maintained.

An ancillary issue to this front are the demands placed on the ethics committees to be sufficiently well grounded in the nature of spatial data and the implications of using different referencing systems. This will have an effect on spatial granularity, and thus raise further issues of confidentiality and subject anonymity. Despite the submission of several illustrations of postcode geographies alongside the proposal to allay fears

of subject identification, it was obvious that a clear set of guidelines are called for. Within these guidelines there should be a series of pre-agreed spatial granularities that are both ethically and technically feasible for dissemination to proposals.

- FRONT 2: The liberation of ethics committees from some of their constitutionally imposed rigidity.

An extension of the issues raised by Front 1 is the major problem associated with obtaining different data sets necessary to complete the analysis exercise. These various data are often judiciously protected by different ethics committees, each working within separate remits. In the current climate, the ability to try and satisfy the legitimate concerns of each committee is an impossible task.

The establishment of an accredited network of organisations with a mutually agreed ethical policy would help the situation. Given that each member organisation has signed a memorandum of cooperation, facilitating the sharing of data in a way that met a strict protocol, the benefits to all participating bodies could be significant. These benefits need not be confined to the network members but - if governed properly - society as a whole, while protecting the identities of the individuals that comprise it.

The foundation of such an accredited network has been tackled in the United Kingdom, to some extent, through the formation of 376 Local Crime and Disorder Reduction Partnerships. This Government led initiative is focused on tackling crime and disorder issues at the local level. Central to its operation is data sharing through a caveat of the Data Protection Act, facilitating the exchange of information between members. However, to date there have been numerous issues regarding the agreement and finalisation of information sharing protocols (see Chapter Four). However, the establishment of these Partnerships still represents progress and should help pave the way to improved mechanisms for data dissemination within a specific network. Membership to a Partnership is via an accreditation schema approved by all parties where ethical issues are the responsibility of the group as a whole rather than individual agencies.

- FRONT 3: Principle that labelling will only be used to ameliorate what is problematic in an area.

The problem of negative labelling has already been stated. However, the potential misuse of a negative label should not be allowed to impede the development of a diagnosis tool that can offer substantial benefits to those most at risk. However, clear rules need to be agreed that enshrines the principle that the label will only be used to ameliorate what is problematic in an area. This will require the appropriate use of the individual data sets and determining legitimate uses for the resulting predictions. This will involve, but not be limited to, a clear statement of whom will see the prediction (in full or aggregate form) and the legitimacy of that viewing.

- FRONT 4: The use of technology to facilitate anonymous “black box” joining and interrogation of data.

One of the legitimate ethical concerns, when considering joining data sets, is the potential for interrogating the composite data set in a way that reveals previously hidden information that might cause harm to the data subject. However, modern technology offers the potential to join data sets in a way that hides the “joins” and only reveals information (carefully chosen and monitored) that is deemed beneficial either to the data subjects or to those associated with the data subjects. A paradigm for how such a system might work is outlined below.

First, the data sets that are to be used in the system would be coded - using a different coding system for each data set - in a way that facilitated them being joined together using criteria for coupling. Second, a “black box” - specifically designed for anonymous data joining and data interrogation - would be fed to the individual coded data sets and “anonymously” joined to form a composite model. Finally, the “black box” interrogates the anonymised composite model to determine the required, previously defined information.

The above approach has the credential of allowing data sets to be joined and interrogated in a way that facilitates the answering of pre-determined questions. For example, the question might be “where are the crime hot-spots likely to be on Saturday?” The ethical implication of answering such a question can be determined before the question is asked. The advantage or disadvantage (depending on point of view) is that the “black box” will answer only questions previously determined. This restricts the full capability of data mining, in that its potential to detect salient trends

in the composite model, and thus answer questions not precisely articulated, is deliberately curtailed. This curtailment of the “black box” capabilities does however ensure that the opportunity of thinking through the ethical implications of answering each question is afforded.

#### **4.6 Summary**

The original research proposal aimed to use data for purposes beyond those for which they were originally collected. This coupled with the requirement for geographical identifiers created unresolvable concerns. Four fronts were presented by which a tentative appeasement of the concerns could take place.

Appeasement of the ethical issues for the continuation of this research was ultimately achieved on several fronts. In one case the perspective discussed by Front 1 was taken forward, the implication being a removal of spatial descriptors in preference to the detailed contextual parameters, while the reverse was used in another instance. A second segment of the research involved accreditation to the Local Crime and Disorder Reduction Partnership, thus working within a pre-defined set of agreed data exchange protocols.

## 5 MODELLING DISPARATE CRIME AND DISORDER DATA

*Modelling crime and disorder through aggregation of multiple data sources can provide insight into community problems and thus help tailor future initiatives. Crime and Disorder Reduction Partnerships (CDRP) have recently been founded as a statutory Government requirement to facilitate data aggregation and analysis at the local level. This Chapter presents HASCADE (Holistic Approach to Strategic Crime And Disorder Evaluation) a non-prescriptive methodology whereby disparate data can be conjointly modelled to direct future policy and instigate targeted initiatives. The research in this Chapter was carried out while working on the audits for two CDRP.*

---

### 5.1 Introduction

Crime and disorder as recorded by the Police, constitutes only a partial descriptor of community issues (Shepherd, *et al.* 1989). Therefore, to understand the dynamics and requirements of a region, there is the need to consult additional data, sourced from a range of organisations at the local level (Graham, *et al.* 1998). On this basis, local partnerships have been promoted to guide and facilitate the data collation, aggregation and analysis process. Hough and Tilley (1998) outline six guiding principles that support the requirement for local partnerships:

- *The police alone cannot control crime and disorder;*
- *No single agency can control crime and disorder;*
- *Agencies with a contribution to reducing crime and disorder need to work in partnership;*
- *Evidence-based problem solving approaches promise the most effective approach to reducing crime and disorder;*
- *Problems of crime and disorder are complex, and there are therefore no panaceas;*
- *Crime and disorder problems need to be understood in their local contexts and strategies need thus to be locally tailored.*

(Hough and Tilley 1998: 1)

The requirement to minimise community problems through tackling crime and disorder issues was formalised through The Crime and Disorder Act (The-Stationery-Office, 1998). The Act places a legal obligation on the local authority and police to work in tandem to develop, publish and implement three-year strategies to tackle crime and disorder based upon findings of a local crime and disorder audit. In addition, the Act stipulates the necessity to work with other key agencies, including health, education, business and voluntary sectors (Sections 5-7, Crime and Disorder Act 1998).

At the time of writing, the local partnerships (CDRP) have recently completed their second three-year audit to direct local strategy until 2005. The aim of the audit is to provide a snapshot of local crime and disorder issues. However, they do not constitute traditional audits in the sense that no attempt was made to achieve a costing of crime, which could be subject to inconsistencies involving numerous arbitrary assumptions (Maguire 1996). That said, some costings do have some merit in terms of quantification of problems (for example, cost to the public and individual of vandalism and car crime). The main objective of the audit was to provide an insight into the scale of crime and disorder over time, commencing with the appraisal of achievements since the previous audit. Central to achieving this objective was an analysis of the region in order to understand the inherently complex nature of crime and disorder dynamics. The Home Office advocates a geographical appraisal. However, a recent report of the auditing process (Phillips 2000) revealed that less than half (42%) made use of a GIS.

The important elements of strategy development rely heavily upon the outcome of the analysis within the audit. *"A model strategy is one that: Is analysis driven. Explicitly understands the process by which the plan is to reduce crime"* (Curtin, *et al.* 2001: 22). The analytical facet of the audit, therefore, plays a crucial role in the strategic development of policy that offers common advantages to partners. To develop a holistic strategy, the analysis of multiple datasets is required, in order to establish a strategy that incorporates the multitude of partner agency objectives. This is necessary to comply with the Crime and Disorder Act (1998). Section 17 of the Act states that *"without prejudice to any other obligation imposed on it, it shall be the duty of each authority...to exercise its various functions with due regard to the likely effect of the exercise of those functions on, and the need to do all that it reasonably can to prevent, crime and disorder in its area"* (The-Stationery-Office 1998): Section 17-1).



Guidance on frameworks and techniques are currently limited. A report by the Home Office (Phillips *et al.* 2000) documents the wide variety of approaches adopted by the 376 CDRP across England and Wales in the previous audit process. This variety provides an indication, arguably, of the limited direction offered. The Home Office toolkits (Home Office 2002) offer some advice, however they fail to develop such guidance into a format that is easily translated for Partnerships implementations. As stated in the toolkit, the aim is one of practical advice offering multiple techniques through which data can be incorporated and analysed. A holistic approach utilising a broad range of multi-agency data is promoted, but techniques by which these can be systemically incorporated are not detailed.

Working with two CDRP (Cardiff and Barry), this chapter presents research that resulted in the identification of a variety of technical and political barriers that currently impede the audit process. Through the use of a non-prescriptive top-down, bottom-up framework some of these issues were alleviated, and a new methodology (HASCADÉ) implemented. HASCADÉ was designed to address some of the deficiencies of current approaches, through a geo-statistical analysis of the audit data. The outputs can then be used to inform future strategy.

## **5.2 The HASCADÉ model**

The HASCADÉ model (Figure 5-1) was developed to address the limited guidance that currently exists for analysing multiple data sets. The focus of HASCADÉ is to indicate some of the potential causes of crime and disorder; this represents a move away from Situational Crime Prevention (SCP) programmes, which are typically advocated for their comparative simplicity (Hough and Tilley 1998).

The HASCADÉ model uses both spatial and statistical techniques to provide an insight into the dynamics of crime, disorder and vulnerability across the partnership region. The final aspect is the formulation of a framework from which the strategy could be evolved.

Targeting finite resources through specific initiatives forms a core requirement of the audit process. A fundamental element of this process was the identification of Priority Geographical Areas (PGA) within which resources could be deployed. However, prior to this audit, decisions on how to identify PGA were limited.

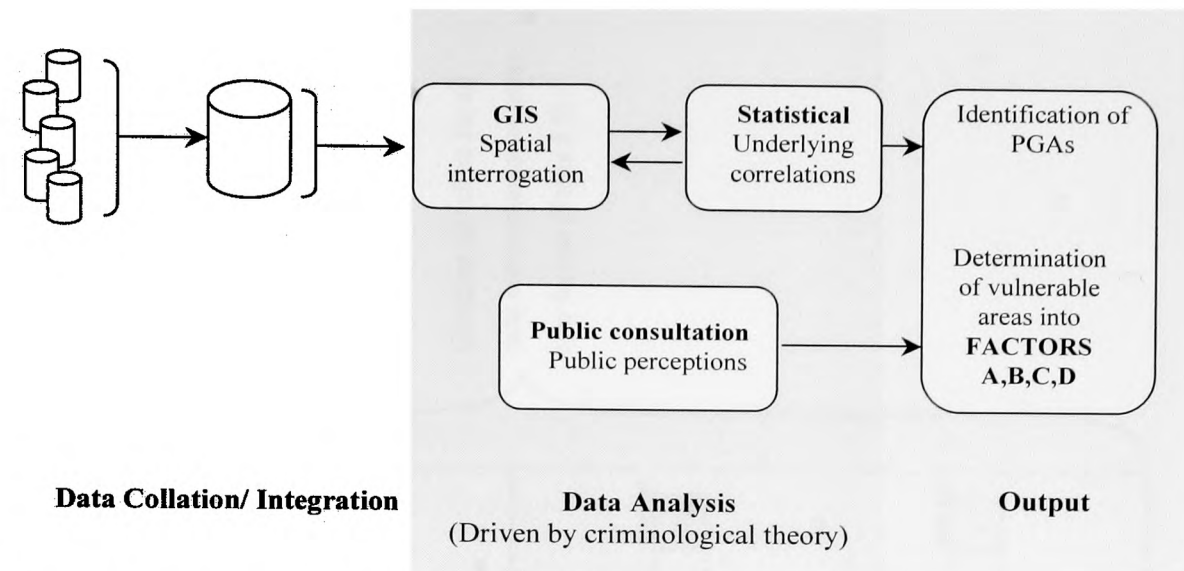


Figure 5-1 The HASCADE model

The HASCADE model introduces a heuristic for identifying PGA on various levels. This was designed to better illustrate the dynamics of crime and vulnerability when considered on such a scale. To achieve this, four factors were used to categorise the variety of issues that formulated the priorities: FACTOR A is considered as a primary priority area; FACTOR B and FACTOR C as secondary, while FACTOR D is regarded as a tertiary priority area (Table 5.1).

FACTOR	FACTOR CHARACTERISTICS	TYPE OF RESPONSE
A	Areas exhibiting several types of crime, disorder <u>plus</u> community vulnerability	A balanced approach: recourse to criminal justice system, situational crime prevention plus prevention of future criminality
B	Area exhibiting one specific type of crime, disorder <u>plus</u> several types of community vulnerability	A balanced approach: recourse to criminal justice system, situational crime prevention plus prevention of future criminality
C	Areas exhibiting several types of crime, disorder <u>but</u> no community vulnerabilities	Recourse to criminal justice system and situational crime prevention methods
D	Areas exhibiting no crime and disorder, <u>but</u> several types of community vulnerabilities	Support to: promote community safety and prevention of criminality. In-depth analysis to highlight any existing good practice

Table 5.1 Factor classification of community vulnerabilities

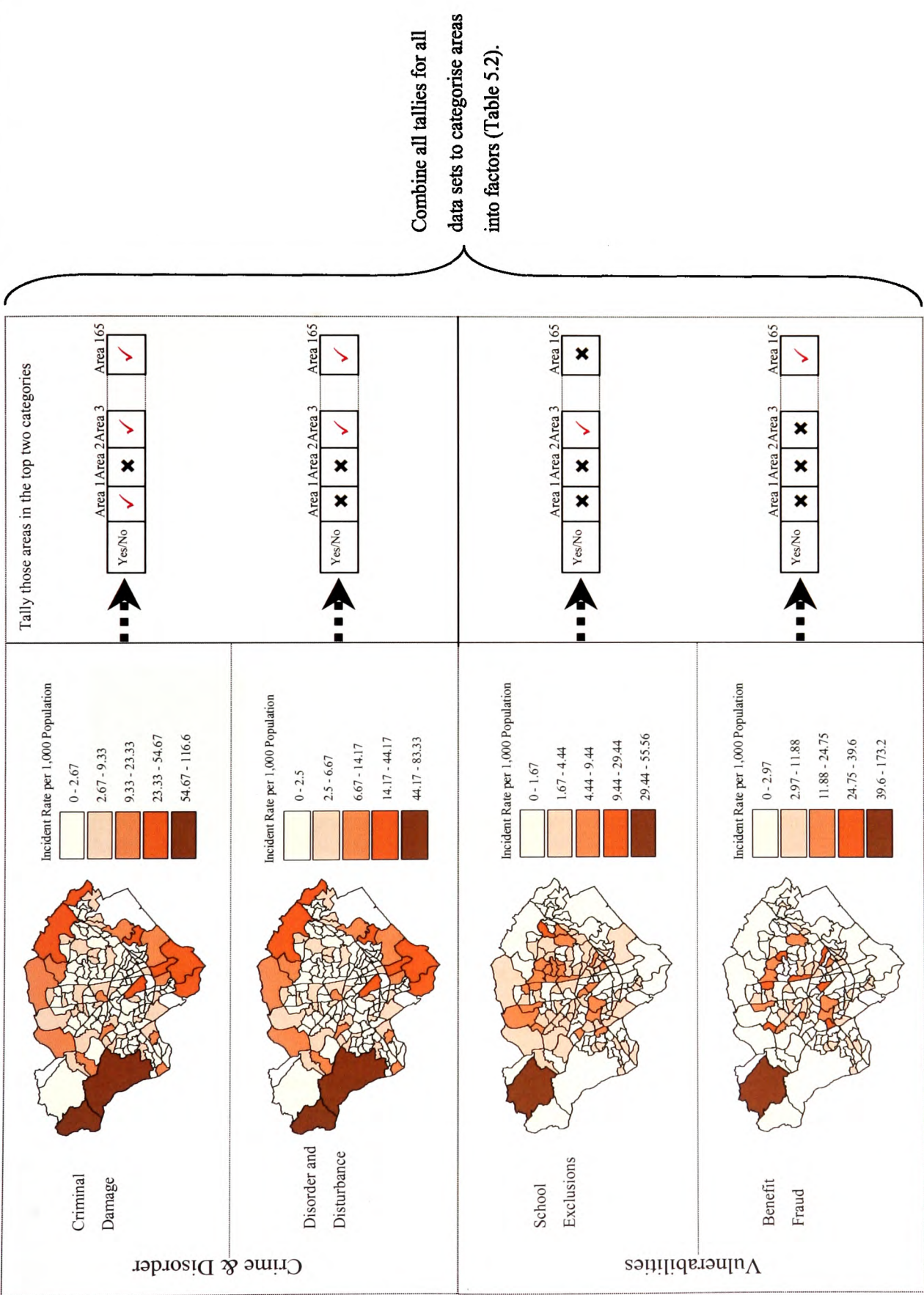


Figure 5-2 The HASCADE count system

To achieve a categorisation of areas using the above factors, a count system was designed. Using a natural break method, areas were spatially represented by five classes. Areas classified in the top two for each data set were tabulated (Figure 5-2). The tabulation of all cartographic outputs was then used to distinguish areas of high crime and vulnerability from those exhibiting lesser levels and achieve the categorisation (Table 5.2). The categorisation was then used to produce a single cartographic output to illustrate their distribution across the CDRP region (Figure 5-4).

CRIME & DISORDER (C&D)		Area1	Area 2	Area 3	Area4	Area165
	Burglary		1	1		1
	Disorder	1		1		
	Violence	1		1		
	Robbery	1		1		
VULNERABILITIES (VUL)	Criminal damage	1		1		
	Benefit fraud	1	1		1	1
	Probation Unsupervised	1	1		1	
	Probation Supervised	1	1		1	
	School exclusions		1			
	YOT Final warning	1			1	
	YOT Police reprimand	1			1	
TOTAL	TOTAL(C&D)	4	1	5	0	1
	TOTAL (VUL)	5	4	0	6	1
FACTOR		A	B	C	D	-

Table 5.2 Classification of small areas into factors.

It was envisaged that by distinguishing between different crime and vulnerabilities within areas, actions aimed at reducing them could be identified, thus providing a holistic approach towards CDRP strategy.

Type of actions to be considered may include:

- Recourse to the Criminal Justice System – such as targeted policing initiatives and increased detection.
- SCP – such as *target hardening* techniques, centred on theories of Broken Windows (Wilson and Kelling 1982) and Routine Activities (Cohen and Felson 1979). Both of

these theories aim to increase the risks associated with committing crime and reduce the rewards.

- Prevention of future criminality – such as youth diversion schemes, increasing educational attainment, increasing access to employment and informing supportive approaches to the allocation of social housing.

### 5.3 Data requirements and issues

The importance of a geographically orientated approach has already been stated. However, modelling the geography of crime in a way applicable to CDRP objectives requires an alternative approach to that demanded by Police operations. Modelling techniques to direct, monitor and evaluate community initiatives demands the adoption of a holistic approach, in which a range of local information is analysed in an appropriate manner.

The ability to visualise the precise locations of events has seen promotion by the Government (Home Office 2000a) and micro-level analysis that has become of particular interest to SCP programmes. Micro level analysis can prove successful in such programmes where the objective is to uncover the specifics of a locale in explanation of its propensity toward observed events (for example a series of houses within a neighbourhood particularly subject to burglary). For the reasons of social inclusion, SCP techniques should not comprise the entire audit analysis. Moreover, in the context of the audit, the use of such techniques places large demands upon each partner to provide full address information from which the data can be geocoded to the fine scale typically demanded by SCP programmes. In addition, the role of the audit is to provide an overview of a whole local authority area, thus fine resolution analysis is arguably not the primary objective. Therefore, the audit should put in place a series of analyses that are capable of identifying the broad issues. At this stage a micro level analysis could take place to isolate the specific issues (for example, vulnerable houses and common modus operandi) to ensure a correct application of preventative measures (for example, a lock fitting scheme and security advice).

During the development of HASCADE the research identified several constraining political, technical and administrative issues. Overcoming these issues required a compromise in terms of data provision in numerous instances. Each are now discussed in turn:

#### 5.3.1 Technical Issues

A fundamental issue at the commencement of auditing was the identification of key technical personnel within each Partner agency possessing the necessary technical

skills and knowledge pertaining to their data systems. This formed a vital stage following which questions regarding data (for example, descriptions of coding protocols) could be posed and replied to efficiently. The most time consuming task prior to the personnel identification were the delays in response to technical questions (a function of being passed around numerous individuals) coupled with the limited knowledge of software systems, data collection and output procedures. Therefore, during the preliminary audit stages, the aforementioned issues contributed to an uncertain data contribution by each Agency.

### 5.3.2 Security Issues

Despite the caveat in the Crime and Disorder Act facilitating the sharing of previously internalised data, many concerns for Partners still remained. This was partially resolved through the implementation of data security measures in addition to the specification of what could be disseminated. Clearly stated dissemination protocols were enforced, whereby no information was to be published (cartographic or text based) beyond the confines of the CDRP without formal approval by all Partners. On a technical level it was agreed that a single machine would be utilised for all data analysis and presentation, the access to which was strictly controlled.

### 5.3.3 Data issues

Many Partners had concerns regarding the sharing of data. The concern was in perceived contravention of the Data Protection Act, despite the clear caveat stipulated within the Act facilitating their use. To appease such concerns, a data wish list proforma was designed to adhere to the general concerns (a function of each data guardian operating within distinct ethical remits – see chapter four), whilst providing the necessary information from which targeted mapping and statistical output could be generated. Using the identified technical contacts all partners completed this and a final standard was agreed, with each agency providing the following as a minimum:

ATTRIBUTE	CONTENT
Temporal Reference	Date (dd/mm/yyyy)
Spatial Reference	Full postcode
Incident or Event description	Numeric code or standardised text
File formats	MS Excel, MS Access, Delimited text

Table 5.3 Minimum data provision proforma

In addition, the confirmation of any problems, inconsistencies and known errors were established prior to the analysis, which in the main involved changes in counting rules and coding protocols. This established known, but not necessarily published, information concerning data reliability. The result of this process was either to omit or amend their use within the audit. This could then be attached to the audit, not necessarily as a formal appendix, but as a reference from which decisions to include or omit certain facets of data could be supported.

One of the most problematic issues was achieving a common temporal coverage across all contributing Agencies. Typical obstacles to accomplishing this were modifications to software systems that rendered data prior to particular dates difficult to access. In addition, the requested data spanned alterations in collection protocols, and thus introduced potentially immeasurable inconsistencies. The result was a reduction in the temporal coverage of all data. A 12-month scale was ultimately agreed to provide the most consistent data log for the whole CDRP region.

#### **5.4 Towards a new approach**

The HASCADE model (Figure 5-1) provides a flexible framework for analysing multiple datasets, including data from a variety of sources. Prior to data being requested, consideration was given to what those data represent. The HASCADE model aims to identify community vulnerabilities as well as crime and disorder events. These community vulnerabilities can provide an insight into the community composition across a region. The following sections present the key techniques and data sources involved in the HASCADE model.

##### **5.4.1 Spatial and statistical techniques**

The analysis of data was conducted through spatial and statistical techniques, utilising Partner data, in addition to the results from a public consultation exercise (discussed later). Central to the entire analysis process was a determination of software availability through which the data could be examined. As a minimum, the analysis demanded use of a GIS and statistical package; these requirements were satisfied through the availability of MapInfo and SPSS. Central to the design of the model was a reliance on standard software functions, commanding minimal customisation, thus reducing the complexity of implementation, use and maintenance. In addition, the model design was founded upon existing criminological research into risk factors and their influence upon vulnerability. These were considered to be the most appropriate for the objectives of the audit. Spatial and statistical techniques were used to identify such risk factors (for example, the number of school exclusions and their



relationship to other data in the area). Results from such analysis could subsequently be used to profile areas in relation to one another, with vulnerable locales identified from those exhibiting lesser levels of risk.

Criminological research (Farrington 2002) has identified a number of factors that when clustered together in an individual's background can lead to an increased risk of them exhibiting future criminality. However, interactions between various risk factors are unclear and numerous studies provide a variety of conclusions as to the nature of such interrelationships. For example, *"adolescents living in physically deteriorated and socially disorganised neighbourhoods disproportionately tend also to come from families with poor parental supervision and erratic parental discipline and tend also to have high impulsivity and low intelligence"* (Farrington 2002: 680). Thus, establishment of causal factors to establish the way each interacts with one another presents an intricate issue to unravel. However, for the purposes of the audit, the necessity to unravel individual explanations was not required. Instead the audit took a broader perspective of risk factors, seeking only an explanation at the neighbourhood level of abstraction.

On the strength of the published evidence, a series of generic risk factors were identified. This process was guided principally by the availability of data. Risk factors (or what can be termed community vulnerabilities) included:

- Economic poverty;
- Peer and/or family criminality;
- Community disorganisation;
- Lack of educational attainment;
- Lack of commitment to school;
- Other social exclusion issues.

The Audit collated information from Partner Agencies to identify areas across Cardiff that are descriptive of a variety of vulnerabilities. The various datasets were considered as representatives of such factors noted above, with the following indicating the vulnerabilities that they represent (Table 5.4).



VULNERABILITY	RELATED DATA SETS
Economic poverty	<ul style="list-style-type: none"> <li>• Council tax benefit claimants</li> </ul>
Peer criminality	<ul style="list-style-type: none"> <li>• Youth Offending Team (Police Reprimand, Final warning, Sentence)</li> <li>• Probation (Supervised and Unsupervised)</li> </ul>
Family criminality	<ul style="list-style-type: none"> <li>• Youth Offending Team (Police Reprimand, Final warning, Sentence)</li> <li>• Probation (Supervised and Unsupervised)</li> </ul>
Lack of commitment to school	<ul style="list-style-type: none"> <li>• School exclusions</li> </ul>
Future risk of lack of educational attainment	<ul style="list-style-type: none"> <li>• School exclusions</li> </ul>
Social exclusion	<ul style="list-style-type: none"> <li>• All of the above</li> <li>• Looked after children</li> <li>• Benefit Fraud</li> </ul>
Crime and disorder	<ul style="list-style-type: none"> <li>• Police incident figures</li> <li>• Results from community safety consultation</li> <li>• Benefit Fraud</li> </ul>

**Table 5.4 Community vulnerabilities and related data sets**

#### 5.4.2 Community safety public consultation

The public consultation (see Appendix D) was designed to further inform the outputs achieved through use of Agency data analysis, assisting in the identification of priorities (locations and issues). The consultation questionnaire was designed to target both district wide anxieties and neighbourhood specific concerns. Through the response to questions, it was possible to assemble neighbourhood level profiles from which comparisons could be made with the recorded Agency data. The two key points addressed were:

- Individuals who had witnessed or been a victim of crime and disorder;
- Perceptions of crime and disorder in the entire area and within the respondent's own neighbourhood.

In an attempt to achieve coverage of all inhabitants, the questionnaire was translated into all recognised community languages (English, Welsh, Bengali, Chinese, Somali, Urdu, Hindi and Punjabi) and distributed throughout the CDRP region.

In total there were 1,816 responses (1,013 were female, 755 were male, and 48 did not disclose their gender). Respondents came from a variety of locations throughout the CDRP region, with the results arguably offering balanced opinion. Key to analysing the outputs from the consultation, and their association with the wider audit data, was the incorporation of the respondent's postcode, through which a spatial appraisal was conducted. Thus, a coupling of respondent's perceptions of their neighbourhoods and city centre in terms of crime, safety and vulnerability together with the Agency derived data analysis provided a powerful tool from which the insight into potential mismatches could be addressed.

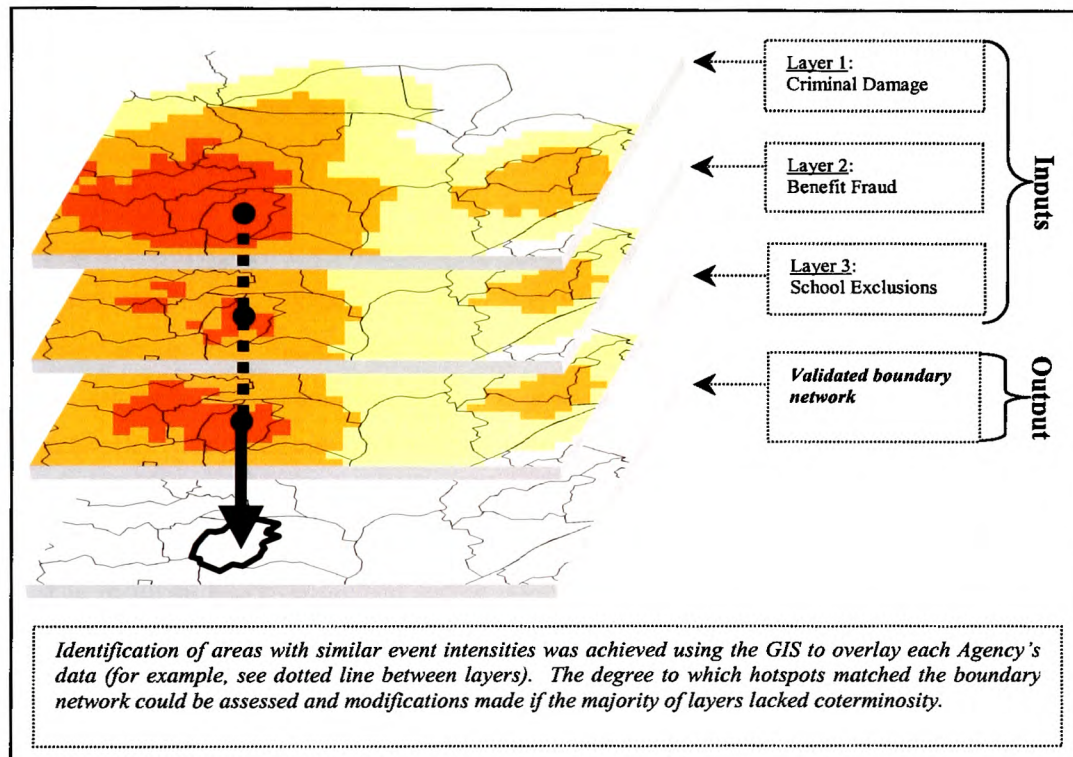
### **5.4.3 Spatial Analysis**

Using the spatial reference provided by all Partners, data was geocoded utilising OS Code-Point® (Ordnance-Survey 2001) to assign x,y co-ordinates at a postcode unit level. Once georeferenced data sets were inspected at point level they were overlain with street and boundary information to provide context. In many cases the volume of mapped incidents created visualisation problems at the point level, where multiple incident localities appeared as a single occurrence. Therefore aggregate and density mapping were used to provide a better indication of event intensity across the CDRP region.

In order to provide a greater context to the underlying population geography to which the various events were related, it was necessary to generate a series of aggregate maps. Using a boundary set based upon the 1991 Census enumeration districts (containing 1999 population estimates) each Partner's data were aggregated to the new framework. A GIS script was created to automate the calculation of number of incidents contained within each region and derivation of rate based upon population. The script consisted of a "point in polygon" test for the Partner's data to calculate the total number of incidents occurring in each region. The total count was then used against the population for that region to derive the incident rate per 1,000 population.

A key part of the spatial analysis was to derive a boundary network to provide the closest representation of each Partner's data. This boundary network then formed the foundation from which statistical analysis could be conducted. Creating and validating the boundary

network involved examining the event distributions (using hotspot mapping - as this best describes incident distribution) from each Agency's data. Where there was an identified lack of coterminosity, the boundary network could be modified to provide a closer fit to the Agency data. Typically, a modification included aggregating two or more regions together. In certain circumstances, however, the imposed boundaries could either under or over fit areas of high event volume. As a perfect match could not be achieved it was deemed acceptable where the boundaries were generally representative of all data, (Figure 5-3).



**Figure 5-3 Method of validating the boundary network**

One limitation of this process was the input boundary network constraining the minimum size of areas for which population data was available. Thus, if Agency data under fitted a region this could not be redefined to a sub-division, as incident rates could not be calculated. A second limitation was the decision criteria used to assess whether areas were representative or required modification. For this stage visual inspection rather than any quantitative techniques were employed as it was recognised that high level precision could not be achieved due to the nature of the Agency's data. A visual comparison was therefore considered sufficient.

At this stage it was possible to commence a primary identification of PGA on the strength of point, hotspot and aggregate mapping across all data sets. Aggregate outputs offered an indication of vulnerability at a broad neighbourhood scale, while point and hotspot maps identify more specific sub-neighbourhood localities internal to these regions.

#### 5.4.4 Statistical Analysis

Spatial analysis, using GIS techniques offered a tool by which visualisation and aggregation was conducted. In many audits, outputs from this stage are taken no further. Therefore one was able to overlay and visualise, but unable to quantify interactions between various layers of information.

Statistical analysis targeted the correlation between the various data sets to reinforce relationships identified through the spatial analyses. Establishing significant statistical linkages, together with identification of PGAs, provided the foundation from which Partnership strategy was developed through building a fundamental comprehension of criminogenic processes (Table 5.5).

Pearson correlation coefficients were used as the basis from which significant relationships were established. Using the rates for each data set, for each small area, significant correlations were flagged and used as the basis to establish key dependencies between the various data areas (small area is used to term each of the areal units of the validated boundary network). Thus, for a neighbourhood area identified as a vulnerable locale for events such as school exclusions and youth offending, spatial inspection can be further queried through statistical linkages, in addition to suggestions of further potential associations that may exist. The result of this hypothetical scenario would be to identify a target set of agencies required to jointly direct interventions in the specified area. In addition, such output and evidence produced through the implementation of this framework reinforces the necessity of the Partnership supporting the current drive towards joined-up government working practices.

Pearson Correlation (2-tailed) n=163		POLICE				LOCAL AUTHORITY				PROBATION				YOT		
		All Incidents	Criminal Damage	Theft and Handling	Benefit Fraud	Council Tax	Looked After Children	Permanent School Exclusions	Supervised Probation	Unsupervised Probation	Police Reprimand	Final Warning	Sentence			
LOCAL AUTHORITY	Benefit Fraud	0.104	0.145	0.082	-	0.276	0.191	0.096	0.436	0.412	0.221	0.100	0.143			
		0.185	0.064	0.299	-	0.000	0.015	0.223	0.000	0.000	0.005	0.204	0.068			
	Council Tax	0.087	0.228	-0.038	0.276	-	0.445	0.322	0.801	0.809	0.303	0.406	0.539			
		0.270	0.003	0.627	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000			
	Permanent School Exclusions	0.074	0.202	-0.017	0.145	0.706	0.482	-	0.639	0.626	0.194	0.381	0.441			
		0.348	0.010	0.831	0.066	0.000	0.000		0.000	0.000	0.013	0.000	0.000			
PROBATION	Supervised Probation	0.112	0.267	0.005	0.436	0.801	0.441	0.22	-	0.852	0.334	0.410	0.425			
		0.154	0.001	0.953	0.000	0.000	0.000	0.001		0.000	0.000	0.000	0.000			
YOT		0.054	0.195	-0.009	0.143	0.539	0.399	0.313	0.425	0.550	0.188	0.297	-			
	Sentence	0.493	0.013	0.911	0.068	0.000	0.000	0.000	0.000	0.000	0.016	0.000				

Table 5.5 An example of the correlation coefficients (significance at 99% denoted by a shaded cells)



### 5.5 HASCADE results

The spatial processes resulted in the production of a single cartographic output (Figure 5-4), supported by the statistical analysis. This final output was designed to minimise the use of numerous representations (for example, graphs, tables and multiple maps), providing a concise indication of crime and vulnerability across the CDRP region, from which strategy could be derived.

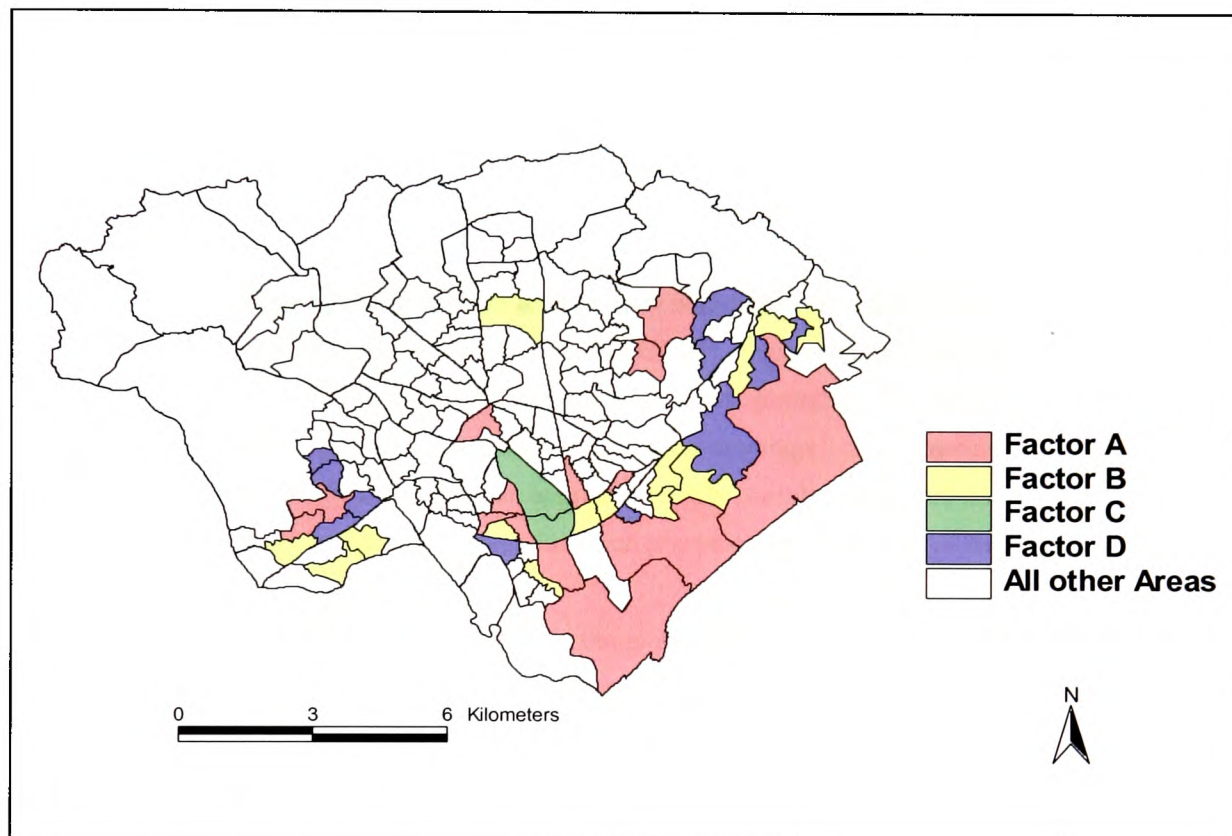


Figure 5-4 Factor map showing PGAs

### 5.6 Future developments

The HASCADE model (Figure 5-1) was the first stage in the implementation of a framework capable of collating, analysing and directing policy to tackle community safety issues. The following section discusses a series of themes by which the HASCADE model, and its role in the audit process, can be further developed.

During the outset of the audit process, a considerable portion of time was taken in developing key technical contacts with Partner Agencies and the agreement of data exchange requests. Future improvements should focus on the formalisation of this process, whereby a register of key personnel are identified (and maintained), alongside meta-data standards of their data. This should be coupled with any published (and unpublished) inconsistencies, errors, coding

and counting rule amendments. Therefore, given a change in audit or Agency personnel, knowledge of the data and systems are not lost.

Dissemination of audit findings to Partners and the general public is of prime importance in promoting the continuation of the CDRPs in their value to community wellbeing, and herein lies the potential for further work. Currently, the majority of dissemination is in the form of printed documents and electronic transmission via web sites for some CDRPs. Web sites in addition to traditional paper based dissemination methods (for example, local newspapers, and leaflets/reports) with appropriately structured information and links to other pertinent sites (for example, police force and the local authority) offer significant opportunities for CDRP operations, strategy and policy promotion (targets and results). In addition, interactive web-based mapping allowing users to interact with community data sets, from which the audit strategy is based, is yet to be established in any CDRP. Such establishment could lead to the CDRP offering a valuable community service and potential for amalgamation with surrounding CDRPs providing superior information coverage. Implementation of such a service would obviously necessitate an appraisal of data confidentiality and issues of negative labelling, thus data content and the way in which it is presented would be of key importance.

The regularity at which audits are currently completed (three years) in many instances may be inappropriate to identify, respond and monitor certain community issues (for example, a burglary initiative). Through the establishment of the data sharing proforma (as a precursor to the HASCADE model), partners are better informed of CDRP requirements and thus better equipped to provide the 'right' data on the first time of asking, thus optimising both time and resources. With this in place, data provision intervals could potentially be shortened to allow analysis through HASCADE and results disseminated to partners and the public through appropriate means (for example, web and paper based methods).

The use of boundaries for defining PGA is a limitation of HASCADE. Some crime, disorder and community safety issues may not conform to such boundaries (for example some virtual communities). A danger is that the aggregated areas used within HASCADE result in an ecological fallacy (Robinson 1950), the misclassification of areas so that they fail to accurately represent the true underlying characteristics. Similar to other methods of analysing information using GIS, outliers will be lost with HASCADE. Such scenarios may require a more detailed statistical appraisal of the data.

The introduction of additional techniques to enhance the current HASCADE model brings with it many difficulties and dangers. The model to date was founded on the use of standard software functionality, thus demanding minimal technical knowledge for its ongoing use and maintenance. However, one future issue could be the focus on crime and disorder over different granularities of time, with the current model identifying patterns over a 12 month period. This could take the form of graph and cartographic outputs, or use of animation to visualise changes over both space and time (see chapter six). Under the currently agreed data provision proforma all Agencies would be able to offer information that could be analysed at a daily level of abstraction. Arguably the use of lower temporal granularities (for example, hourly) would present far greater demands on the analyst and is currently beyond the strategic overview required by the audit process.

### **5.7 Summary and Conclusions**

Currently, direction of the audit process is subject to limited guidance and open to interpretation by each of the 376 CDRPs. The HASCADE model was developed to offer a flexible framework, utilising an amalgamation of spatial and statistical methods embedded within criminological theory, to address community safety issues. Using standard software tools the model was successfully implemented and utilised to analyse a snapshot of crime and disorder, and provided the foundation from which targeted strategy was developed. HASCADE was not dependent upon high levels of customisation. The result was the production of a sustainable, replicable and flexible crime and disorder reduction model.

The model offered a move away from the more traditional approaches, where interventions are targeted at areas exhibiting high crime and disorder levels. This is based upon the underlying assumption of a correlation between high crime and disorder areas with high vulnerability. HASCADE showed that this was not always true, by offering explanations for intervention based upon multiple datasets. The HASCADE model assisted with promoting and supporting the mainstreaming of community safety, as both statutory and non-statutory partners were incorporated into the process. The strategy document (Bowen Thomson and Corcoran 2002) details programmes of work for Partners' Agencies over the ensuing three-year period. Three overarching themes were identified: concern of crime; substance misuse and youth issues; and key crime and disorder issues (burglary, vehicle crime, criminal damage, anti social behaviour and violent, hate crime) using the identified PGAs. These themes will then form the foci for targeted initiatives.



Access and use of the disparate data sources detailed in this Chapter reinforces findings presented in Chapters Three and Four. Despite the clear caveat in the Data Protection Act, data provision remained a problematic issue for some contributing Agencies and one that is likely to persist. One potential solution to this was implemented by HASCADE by minimising the level of detail (i.e. spatial, temporal and contextual information) required by the model. GIS was a new requirement for many agencies - data not being readily available in a typically requested format (i.e. full address), and non compliant with the British addressing standards (for example, BS7666). This can be true of voluntary agencies, such as social landlords and domestic violence groups. Using postcodes reduced the reliance upon each agency to collect pin-point (full address) level information, therefore a larger variety of data could be analysed.

## 6 EXPLORATION OF INCIDENT AND INJURY DATA

*Using two data sets procured from the Police and Accident & Emergency (A&E) services, this Chapter presents the results from a variety of exploratory techniques for the identification of patterning and trends. The Chapter concludes with a summary of the key findings and a comparison of the two data sets focusing upon similarities and disparities of incident volumes between each.*

---

### 6.1 Introduction

Crime is a complex phenomena and it would be unrealistic to assume a simple relationship (for example, time of day or day of week alone explaining the volume of violent crimes) as a robust predictor. In order to formulate an understanding of the dynamic nature of such activities, it was necessary to conduct a variety of spatial, temporal, spatio-temporal and statistical deconstructions. The exploratory techniques detailed provided a fundamental comprehension of data interactions from which elementary hypotheses were developed. The hypotheses were then used as a foundation from which predictive techniques would be validated (as detailed in Chapter Seven).

Using data procured from both A&E services and the police this Chapter describes procedures that can be employed to analyse the dynamics of crime and injury within the Cardiff study region.

### 6.2 Data procurement

The data collection exercise was designed to provide a detailed spatio-temporal log of the study area. The log constitutes a series of linked databases, which are descriptive of a range of crime and disorder, socio-economic, holiday, meteorological and public events based information that affect the Cardiff region. The objective of the log was to provide a series of key information, which following analysis was capable of substantiating spatio-temporal occurrence of crime and disorder through the derivation of a model.

As Chapter Four explained, this work was subject to a series of ethical and technical constraints within which the research had to operate. The satisfaction of issues resulted in a reduction of descriptive information, which pertain in particular to criminal activity. Therefore, the inclusion of descriptor variables, such as age, gender and ethnic origin, for example, were unavailable for inclusion in the data. In the case of the A&E data, ethical constraints necessitated the removal of any spatial reference, leaving only temporal and descriptor information.

In addition, issues of data availability, coverage, and consistency affected the procurement exercise. One particular instance pertained to software connectivity and compatibility. In this case the two systems in use were unable to cross-reference data between one another and therefore were incapable of aggregating certain descriptive information. Other problems included the availability of a consistent data source, for example, meteorological information, where specific sensors (i.e. rainfall and temperature) were inactive, thus providing no data.

The result of the procurement process was to produce a “desire versus potential inventory” (similar to the desire wish list discussed in Chapter Five). At the outset a temporal coverage of five years was sought, which was achieved for A&E data and contextual sources, but was reduced to two years for the police data due to various constraints. Both time spans, however, still provided sufficient volume of data in order to model criminal activity. Details of each data source are described in Table 6-1.

### **6.3 Data Inspection**

The first task prior to any form of analysis was a thorough inspection of the core data in its raw form to uncover any potential erroneous elements.

#### **6.3.1 Erroneous and missing data**

An additional component of the exploratory process was to identify all potentially erroneous data. The quality of data is dependent upon a series of events: the type of data; how it was collected and entered; how and when it was edited; and, finally, how data is presented, stored, and intended for end users. As data quality is affected during every part of the collection, structuring and presentation process, data quality forms an integral part of what must be a total quality concept.

As the techniques described here involved the association and integration of a variety of data sources (both spatial and attribute based), and originating from numerous contributors, many potential problems arose in terms of the maintenance of an acceptable level of data quality. It

was therefore of great importance that there was an appreciation of each data set used, inclusive of its potential weaknesses, where feasible (a point previously discussed in Chapter Four).

	Data set	Attributes	Description
Crime	Accident and Emergency	Age, gender, date of attendance	Five-year coverage of assaults cases admitted to Accident and Emergency. <i>All fields complete</i>
	Police Incidents	Date/time, incident type, location (x and y co-ordinates)	Two-year coverage of calls for police service <i>5% missing spatial reference, 0.4% incident type missing, therefore incompatible for inclusion to the analysis</i>
Contextual	Meteorological	Temperature, rainfall, sunshine	Hourly and daily weather data derived from the local weather centre (central Cardiff).
	Sporting/Events	Date/time, sport/event type, attendance	A comprehensive coverage of major sporting and events scheduled in the city detailing type, nature of the event in addition to attendance numbers where available.
	Holiday	Start/end date, holiday type	Five-year coverage of all public and school holiday periods.
	Socio-economic	Location (enumeration district), lifestyle class	A subset of a national data set providing a hierarchical lifestyle classification of the study area at the enumeration district resolution, known as Super Profiles (Brown and Batey 1994) based on the 1991 Census.

Table 6-1 Data sets used in the analysis

Three main categories concerning the accuracy of spatial data sets can be identified, namely:

1. Spatial;
2. Temporal;
3. Attribute and statistical.

Unlike the contextual data sets described above, where precision can be measured more readily, accuracy of incident and injury data is a much less precise science and more difficult to define in terms of how close to reality the data portrays. The data quality problem in essence can be described as the difference between reported incidents and actual incident levels. In the case of policing data, the difference between these two parameters occurs as a result of the fact that not all incidents that occur are either reported by the public, or recorded by the police. In addition, computerised recording systems and georeferencing gazetteers may not be sufficiently developed to accurately record all reported incidents. In many instances systems may fail to adhere to British addressing standards (BS7666) (for example, the Cardiff A&E department, where the initial database design requirements demanded no geographic outputs). Such systems were primarily designed to cater for relatively simple aggregate statistics to be derived for administrative reports from the central data store. Location of incident and victim and perpetrator data is typically entered as text fields where no cross validation against an address gazetteer is available. Thus, if mapping is to be conducted, significant time and resources are typically required to “cleanse” and geocode all records.

Other potential problems associated with the police data collection methodology involve the recording process (i.e. categorisation and assignment of spatial reference), and changes in reporting directives aimed at increased detection rates of certain types (Povey, *et al.* 1997). In addition, previous studies by Bottomley and Coleman (1976), and McCabe and Sutcliffe (1978) have highlighted the implications of individual police officers, and their impact upon the aggregate output statistics. However, quantification and identification of this potentially erroneous element was not possible, and, following personal communications and advice in relation to police data, a series of specific incident types were selected in an attempt to achieve the highest consistency and accuracy throughout the data.

Classification of an assault by A&E staff may also introduce inconsistencies, as individual perceptions will impact upon the overall precision of the data set. This, however, appeared less pronounced to those of the police data by the nature of the direct reporting process (classification of an assault is on the basis of communications with the triage nurse), potentially introducing less errors (for more detailed information on the reporting process see Sivarajasingham and Shepherd 2001).

A partial outcome of these tasks was the amendment or omission of erroneous entries, where deemed necessary. For example, one task, in the case of police data, simply involved plotting

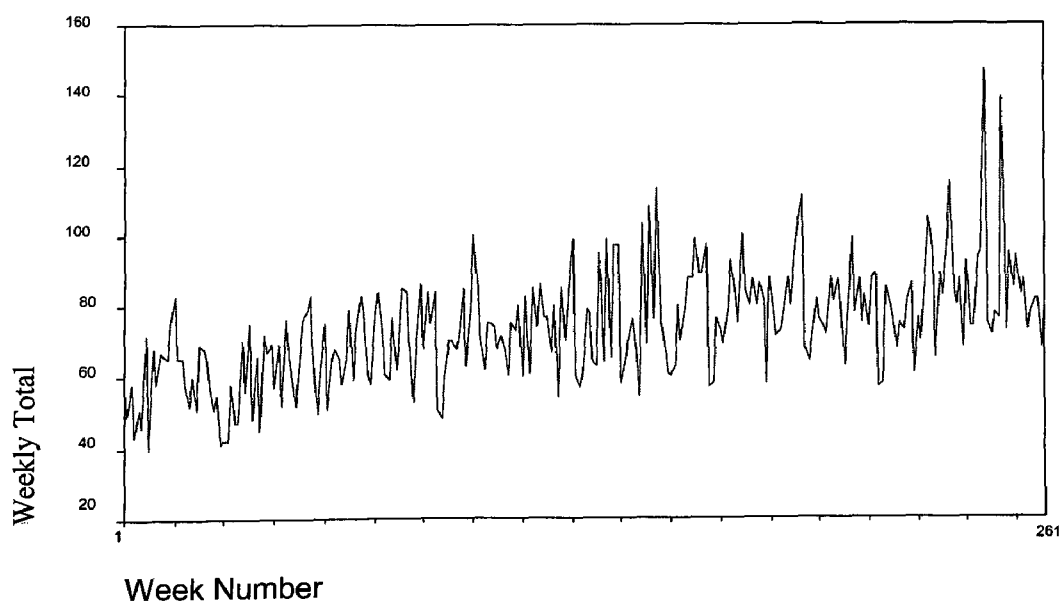
incidents over base mapping (Ordnance Survey raster tiles and boundary data) and omission of those not conforming to feasible locales – a function of incomplete coordinates. In addition, all such entries were logged for future analysis. This process provided supplementary information to that obtained through communications with the individual data suppliers - where known (but not necessarily published) errors were sourced.

The following sections present the results of a series of exploratory analyses for each of the primary data sets (A&E and police).

### 6.3.2 A&E

Use of A&E assault data affords an opportunity to investigate the nature of violence resulting in injury not always recorded in police records (Shepherd, *et al.* 1989). In addition the data permits investigation into the influence of both age and gender, thus facilitating assessment of differences in vulnerability existing between the various groupings. Research utilising A&E data to identify patterns of community violence is extremely limited. However, Sivarajasingham and Shepherd (2001) have conducted a large scale (monthly) analysis of national trends. In contrast, the study presented here focused on the identification of small-scale trends (daily as opposed to monthly) and their effect across age and gender.

A total of five years A&E assault data were analysed (May 1995-April 2000), comprising 19,264 cases, each including age, gender and date of attendance. Over the five year period there was a steady increase (on average 2.1 per month) in assault attendees, coupled with an increase in variability (Figure 6-1). The mean rate was 10.5 per day (73.8 per week).



**Figure 6-1** Total assault cases by week

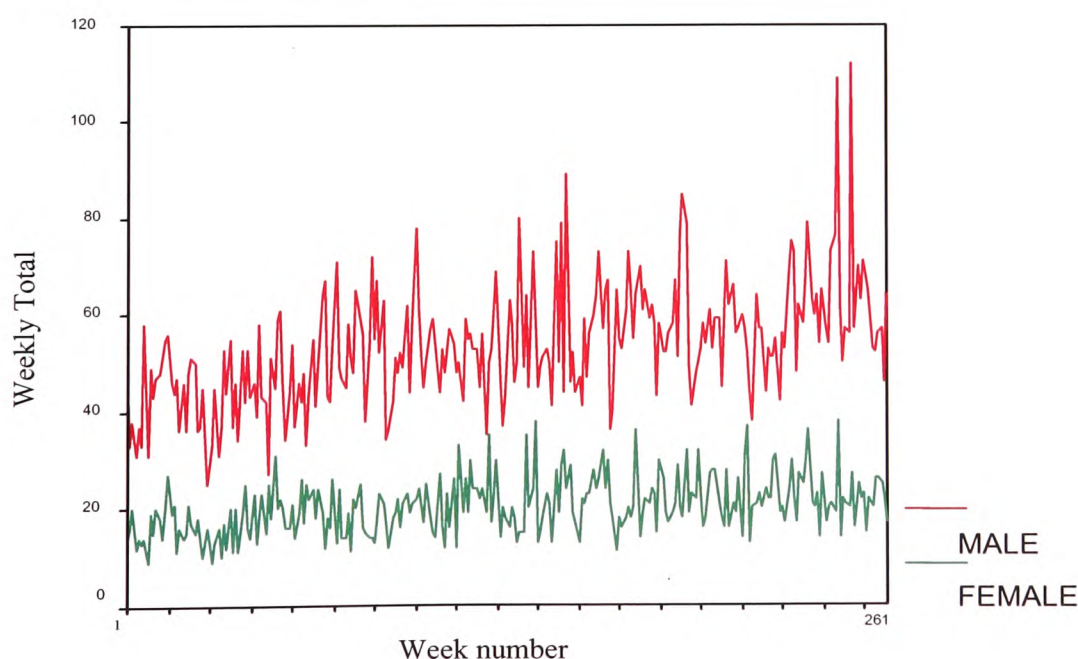
The day of the week was found to be a distinctive factor in the determination of assault rate (Table 6-2). Almost 40% of all cases occurred on either a Saturday or Sunday. Comparing the first and last 52-week period, there was a trend towards a more noticeable peak during the weekend. It was the strong influence of day on volume that focused attention on daily fluctuations, in addition to any discernible trends at weekly, monthly and seasonal aggregations.

	Mean	% of Total
Monday	9.3	12.6
Tuesday	8.0	10.9
Wednesday	8.5	11.5
Thursday	8.8	12.0
Friday	10.6	14.4
Saturday	14.5	19.7
Sunday	14.0	19.0

**Table 6-2 Breakdown of assault by day**

- **Age and Gender**

The gender distribution was very skewed (72% male), Figure 6-2. However the weekend peaks were evident for both male and female attendees, with the patterns being fairly similar.



**Figure 6-2 Breakdown of assault by gender**

The age distribution showed that (approximately) 25% are aged 18 or less, 25% between 18 and 24, 25% between 24 and 34 and 25% are aged over 34. For comparative purposes, the same age groupings as used in the Sivarajasingham and Shepherd study were used for all analyses. Table 6-3 shows the distribution of patients by age and gender.

		Sex		Total
		Female	Male	
0 to 10	Count	208	471	679
	% within Age group	30.6%	69.4%	
11 to 17	Count	909	2260	3169
	% within Age group	28.7%	71.3%	
18 to 30	Count	2204	6894	9098
	% within Age group	24.2%	75.8%	
31 to 50	Count	1663	3725	5388
	% within Age group	30.9%	69.1%	
Over 50	Count	326	604	930
	% within Age group	35.1%	64.9%	
Total	Count	5310	13954	19264
	% within Age group	27.6%	72.4%	

Table 6-3 Distribution by age and gender

Examination at the weekly level once again showed the prevalence of weekends, particularly for the 18-30 age category (Table 6-4).

	0 to 10	11 to 17	18 to 30	31 to 50	Over 50
Monday	15.0%	14.5%	11.1%	13.8%	11.9%
Tuesday	11.5%	12.7%	9.7%	11.0%	14.7%
Wednesday	13.0%	14.9%	9.8%	11.7%	14.5%
Thursday	13.7%	12.9%	12.1%	11.3%	10.2%
Friday	14.0%	16.2%	14.4%	13.4%	15.6%
Saturday	15.3%	16.0%	21.5%	19.8%	16.8%
Sunday	17.5%	13.0%	21.4%	18.9%	16.2%

Table 6-4 Age group by day

A comparison between the first and the last 52 week periods showed overall increases in both male and female attendees (40.4% and 37.0% respectively), with this increase being



most noticeable at weekends (rising by 65.9% for males and 68.0% for females). In particular, this tendency was most pronounced in the 18-30 age category where females saw a rise of 82.4% and males increasing by 69.5%.

Despite the marked increase in the number of females aged 18-30 assaulted during the weekend, they still accounted for only 11.3% of the total intake, while males of the same category comprised 43.5%. In addition, there was a marked increase in the number of both males and females aged 31-50 assaulted during weekends (rising by 86.5% and 70.3% respectively). The volume of cases within the 31-50 age range remained at over 70% of the cases experienced in the female 18-30 category, whilst for males the relative frequency remained below 50%.

#### • Seasonality

To provide a direct comparison to the Sivarajasingham and Shepherd study, three monthly aggregates were used to ascertain evidence of seasonality. Figure 6-3 shows the three monthly aggregates, for the spring (March, April and May), summer (June, July and August), autumn (September, October and November) and winter (December, January and February).

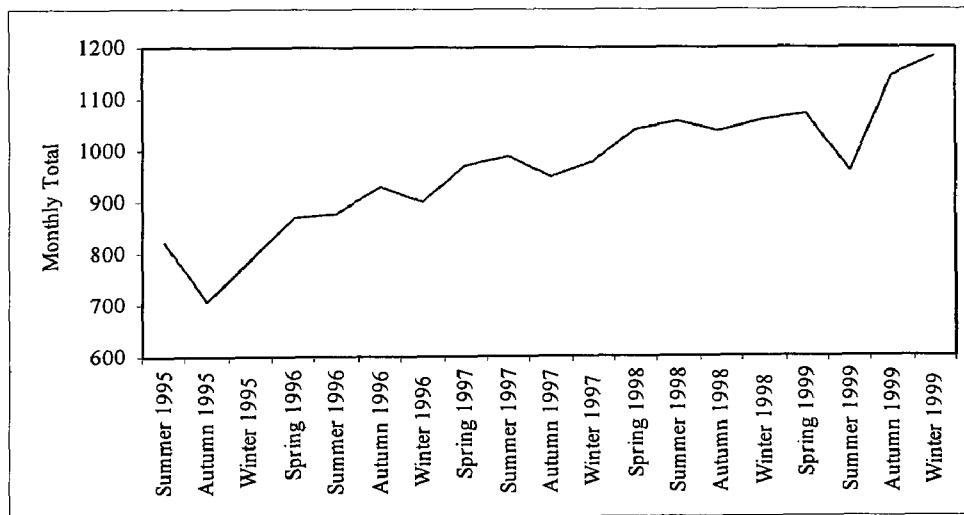


Figure 6-3 Three monthly totals

If a regular cyclical pattern together with a linear growth trend was evident, then this could be identified through the differences between the seasonal totals (Table 6-5).

	Spring Winter	Summer Spring	Autumn Summer	Winter Autumn
1995			-116	+79
1996	+85	+7	+51	-28
1997	+71	+18	-40	+27
1998	+66	+15	-21	+23
1999	+13	-113	+184	+38

Table 6-5 Difference between seasonal totals

Analysis by the previous national study (Sivarajasingham and Shepherd 2001) had highlighted the existence of seasonality, with the incidence of assault being highest during July to September and lowest from February to April. Examination of the Cardiff data indicated no clear patterning over the same time period. Seasonal variation across age and gender was also investigated, while again no apparent patterning was revealed.

Seasonality is well documented in the literature as forming a key component in the rationalisation of criminal activity. It was as far back as the Roman period that the linkage between the level of criminal activity and environmental factors such as temperature was first established (Harries, *et al.* 1984). Since this period, numerous subsequent studies have supported this relationship between temperature and aggression (Guerry 1833; Harries, *et al.* 1984; Anderson 1987; 1989; Cheatwood 1995). A common finding throughout was that increases in the ambient temperature led to increases in aggression and violent crimes. Analysis of the Cardiff data against that of weather (temperature, rainfall and sunshine) first involved the removal of the upward trend. Taking the daily averages over the five years, each were combined to form a series of 60 monthly estimates. These estimates were then subtracted from the original data to create a time series of deviations from the monthly average. The time series was then fitted with a linear trend line, and taking the residuals from this process, correlations between the meteorological variables and assault intake were calculated (Table 6-6).

	Unstandardised residuals
<b>Temperature</b>	0.070 (0.596)
<b>Rainfall</b>	0.156 (0.234)
<b>Sunshine</b>	-0.036 (0.784)

Table 6-6 Pearson correlation coefficients for temperature, rainfall and sunshine against assault intake (significance level in parentheses)

Analysis of the Cardiff data against the weather variables revealed neither a clear seasonal pattern nor any significant meteorological influences explaining the incidence of assaults. Over the five years, the months with the highest assault figures were December and August, which makes a valid explanation based purely on meteorological factors very unlikely.

- **Sporting events and holiday periods**

Calendar effects (Harries, *et al.* 1984; Hylleberg 1995) have also been used to explain seasonal variations of crime. Calendar events describe particular points and periods in the year; public or school holidays, pay dates and sports events for example. Each has the potential to describe seasonal variations in criminal activity, where a fluctuation in a particular type of crime is a function of a temporal occurrence (for example, burglary rates in student rented accommodation in relation to university holiday periods). Prior research has concentrated on the use of police data to derive such relationships. For this research A&E data were analysed to search for similar associations.

Number of cases	Day	Date	Calendar event
62	Saturday	1 January 2000	New Year
46	Friday	1 January 1999	New Year
42	Saturday	5 February 2000	Rugby International
34	Sunday	6 February 2000	"
27	Saturday	23 October 1999	Rugby International
37	Sunday	24 October 1999	"
33	Thursday	1 January 1998	New Year
30	Saturday	18 December 1999	Unknown
29	Wednesday	1 January 1997	New Year
29	Saturday	11 March 2000	Unknown
28	Saturday	15 March 1997	Rugby International
27	Saturday	21 March 1998	Unknown

Table 6-7 Largest A&E attendance's

Analysis of extreme results (Table 6-7) indicated the association of both sporting events (specifically rugby internationals) and the New Year period. Out of the ten busiest days for A&E assault admissions, four out of the five New Year periods are included, with each year progressively busier (113.7% increase from 1997 to 2000). Other holiday

periods (school and bank holidays) and events (e.g. concerts) were examined, but no well-defined relationship existed between any age grouping.

- **Discussion of A&E analysis**

Examination of the 2000 British Crime Survey (BCS) (Kershaw, *et al.* 2000) (survey period approximately coterminous with that of the A&E data coverage) corroborates some findings of the A&E data, indicating that violent incidents tend toward the evening and night, with 43% occurring over weekends (40% A&E). In addition, males aged 16-24 were identified as those most at risk of violence by the BCS, a similar finding to that shown by A&E data. The BCS, however, indicated an overall fall of 4% on the numbers of violent crimes between the period 1997 to 1999 (common assault -3% and wounding -11%). Cardiff figures therefore failed to concur with national trends, with an increasing trend of 2.1 per month identified over the five-year period.

Knowledge of the environments in which the assaults occurred is of value in terms of implementation of preventative measures. The BCS indicated the home and surrounding area as the most likely arena for violence (26%), proceeded by public areas and streets (23%) and pub/club localities (19%). The work of (Shepherd, *et al.* 1993) corroborates this prevalence of public places as arenas for the majority of male assaults. Given ethical approval, this aspect could have been geographically and empirically tested as part of this research.

With the statistically significant decrease in domestic violence of 23% between 1995 and 1999 noted by the 2000 BCS (Kershaw, *et al.* 2000), this study noted increases in female violence within the 18-30 and 31-50 categories. The Sivarajasingham and Shepherd study suggests public places as the probable arena for this violence. Examination of daily patterns by the research presented here highlighted increased tendency toward violence during weekends for both males and females. This in turn suggests that both males and females of such age groups pursue similar activities, leading to similar situations, motivations and thus vulnerabilities to assault.

The previous study investigated trends evident at monthly and tri-monthly aggregations, indicating significant evidence of seasonality. Analysis of the Cardiff data failed to establish such relationships, instead highlighting the importance of weekly cycles of assault. However, despite using smaller time aggregates (daily volumes), the magnitude of certain daily patterns may still have been obscured. Given the availability of data,

subsequent analysis should focus on inspection at an hourly interval, as it was found that although numerous incidents occur on Sundays, it is extremely likely that a large proportion of these are admissions in the early hours (Shepherd 1990). This arguably should be reaggregated to form part of the Saturday figures.

### 6.3.3 Police

Unlike the A&E data, the police data included the location (x, y coordinate) and hour of incidents (time to the nearest hour), but did not contain any additional variables. The focus of the exploratory task thus targeted interrogation of spatial and temporal parameters. The analyses were dissolved into a series of four key areas: spatial, temporal, spatio-temporal and contextual inspection, each with their distinct procedures.

The focus of this research targeted incidents of violence and disorder, as these constituted a large proportion of the total number and offered the possibility for comparison with the A&E data. It was following advice through personal communication that a final series of incident types were selected on the basis of consistency throughout the data coverage period. These included:

- Violence against the person (c10);
- Criminal Damage (c60);
- Disorder/Disturbance/Nuisance (d10; d90).

For a more detailed description of each offence category see (Home-Office 2000)

A total of two years of calls for police service data were analysed (July 1999-June 2001). This comprised 39,125 cases, detailing location, date and time of the incident. Over the two year period there were decreases in disorder and nuisance incidents (d10 and d90) over an initial three month period, after which all exhibited relative consistency (Figure 6-4). Due to the limited timespan of data, seasonality could not be properly addressed.

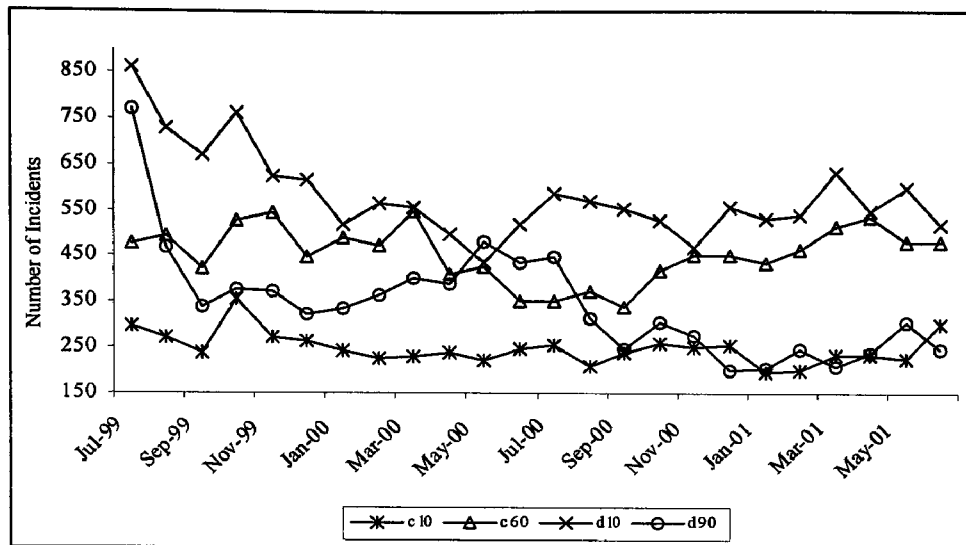


Figure 6-4 Total incidents by month

The following presents the results from each of the four key areas, discussing the techniques employed and concluding with their key findings.

- **Spatial**

Exploratory spatial analysis of the data was performed through visual inspection using pin, aggregate and density mapping techniques, accompanied where necessary with confirmatory statistics (for example, the Moran I statistic testing for global spatial autocorrelation). Pin mapping provided a useful insight at the large scale (street level). However, at smaller scales this approach suffers from the problem of not conveying true volumes and intensities when multiple events occurred at a single locality. In such instances, other techniques, such as variable symbol size or density mapping techniques, were used to better indicate incident volume at the city level. Another technique involved aggregating the point data to enumeration districts and calculating population incident rates to provide an indication of incident relative to the resident population in which they were committed. This provided a useful insight regarding the vulnerability of residential neighbourhoods. However, the results had a tendency to be skewed when including largely commercial or industrial areas (for example, the city centre) where residential volumes are minimal. Hence, these were treated independently.

The Moran I statistic was used in conjunction with mapping outputs to quantify and confirm any observed patterning (Table 6-8) at the global level. Its calculation used an adjustment for small distances constraining the weighting of points in close proximity to one another, thus preventing the I statistic from becoming markedly large (Levine 1999).

This was found to be particularly pertinent to incident data, where urban morphology act to confine the localities of criminal acts. An additional requirement for both tests was the inclusion of an intensity variable (incident volume), whereby observations at one location could be compared to that of another. Following the methodology used by Levine (1999), the enumeration district boundaries were used to derive the intensity variable through aggregation of each incident to their respective district, and assigning the total number as the intensity.

Incident type	Volume		Rate	
	Moran I (Expected I)	Z (Normality)	Moran I (Expected I)	Z (Normality)
d10	0.009 (-0.0017)	5.71	0.007 (-0.0017)	4.38
d90	0.004 (-0.0017)	3.16	0.002 (-0.0017)	2.22
c60	0.017 (-0.0017)	9.90	0.011 (-0.0017)	6.57
c10	0.001 (-0.0017)	1.63	0.0001 (-0.0017)	0.95

**Table 6-8 Results of Moran I for each incident type**

Evidence of spatial autocorrelation at the global level was found to be limited in the direction of a positive autocorrelation – areas of high incidence typically adjacent to other areas of high incidence and similarly low incidence adjacent to low incidence. The intensity variable was calculated using both incident volume and incident rate (per area) neither of which resulted in significant results. One solution could be to utilise a more advanced measure of proximity in order to account for second, third and higher order neighbour zones in addition to those that are directly adjacent. In addition, various zone designs were tested (for example, Census wards, enumeration districts and grids of varying size); the more significant results were derived using the Census enumeration districts and incident volume as the intensity variable.

The use of kernel density mapping techniques offered the possibility to investigate local evidence of clustering that may be less evident in the results of global tests such as Moran.

Key Findings
<ul style="list-style-type: none"> <li>Cardiff encompasses an area 242,700,000m<sup>2</sup>. Given this, it can be said that, on average, one incident took place during the year per 13120m<sup>2</sup>, or approximately one incident per 65m radius</li> </ul>
<ul style="list-style-type: none"> <li>Limited evidence of spatial autocorrelation using the Moran's I statistic. Criminal damage (c60) showed the greatest and violence against the person (c10) showed the least evidence of global spatial autocorrelation.</li> </ul>
<ul style="list-style-type: none"> <li>Kernel density mapping showed c10 incidents to be most concentrated.</li> </ul>

### • Temporal

Time frequency plots (Figure 6-5 and Figure 6-6) provided an indication of incident distribution and trends over various granularities of time, (i.e. month, week, day and hour). Monthly and weekly aggregates of time offered some insight into existence of seasonality (limited by the two year timespan), while shorter intervals highlighted the more variable nature of criminal activity and injury by hour and day.

Over the two year period there was a decrease in disorders (on average -22.8 for d90, -14.9 for d10) with minor increases in c10 and c60 offences (both on average +0.1 per month). The mean rate was 8.1 for c10, 14.9 for c60, 19.1 for d10 and 11.3 for d90 per day (56.9, 104.4, 133.74 and 79.3 per week).

Key Findings
<ul style="list-style-type: none"> <li>Monthly patterns were occluded to some extent by the two year timespan. C10 showed increases during the summer months (June +10.25% and July +11.45% above the mean). C60 in contrast showed elevated levels during January, February and March.</li> </ul>
<ul style="list-style-type: none"> <li>Daily patterns showed very distinctive trends, with significantly increased volumes during weekend days, the lowest point early to midweek. Patterns were similar for each incident type.</li> </ul>
<ul style="list-style-type: none"> <li>Hourly patterns showed distinctive increases towards the midnight hours, especially during weekend days. Lowest volumes were experienced during the early hours. Patterns were similar across each incident type.</li> </ul>



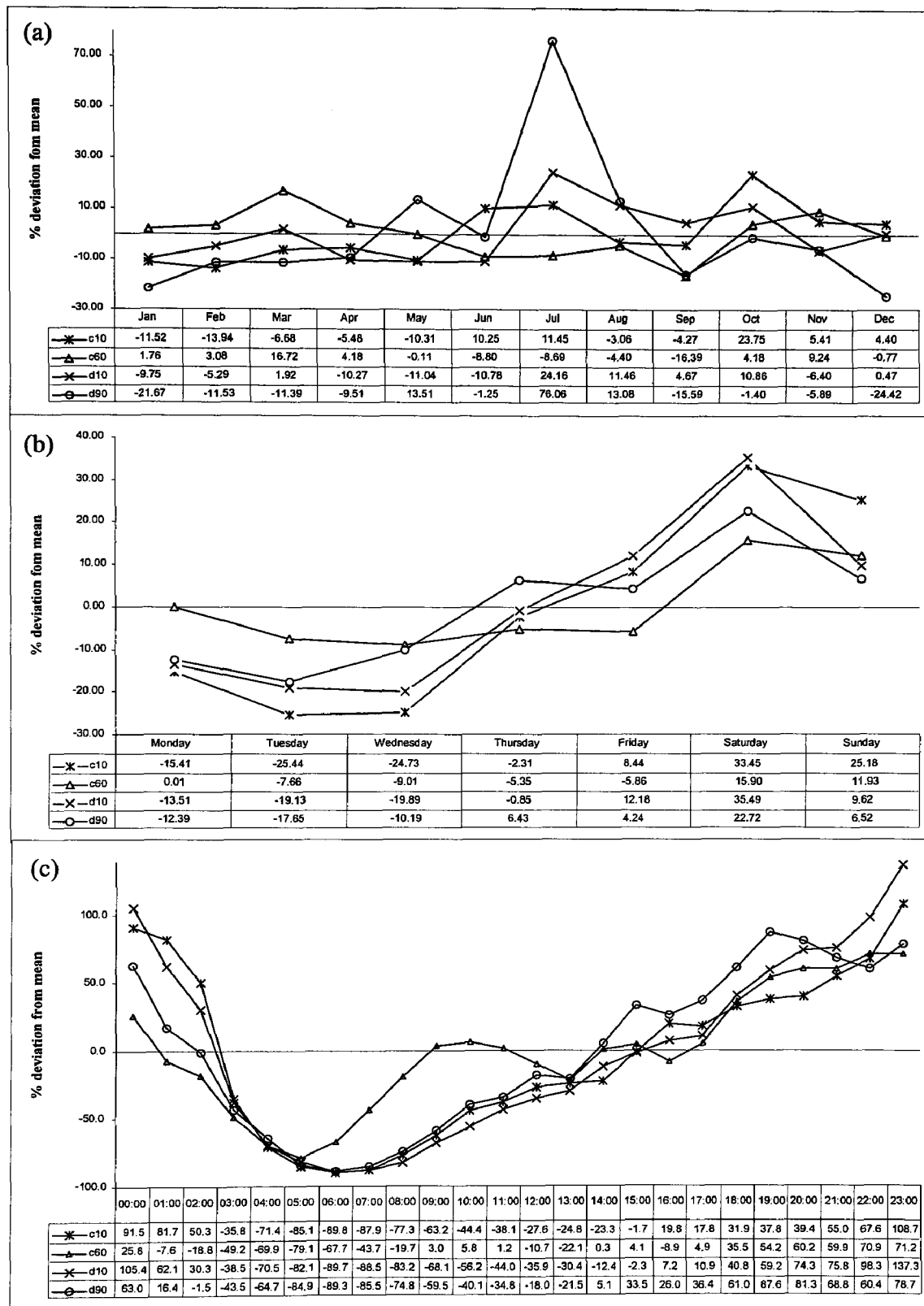


Figure 6-5 Volumes of incident (a) monthly, (b) weekly and (c) hourly.

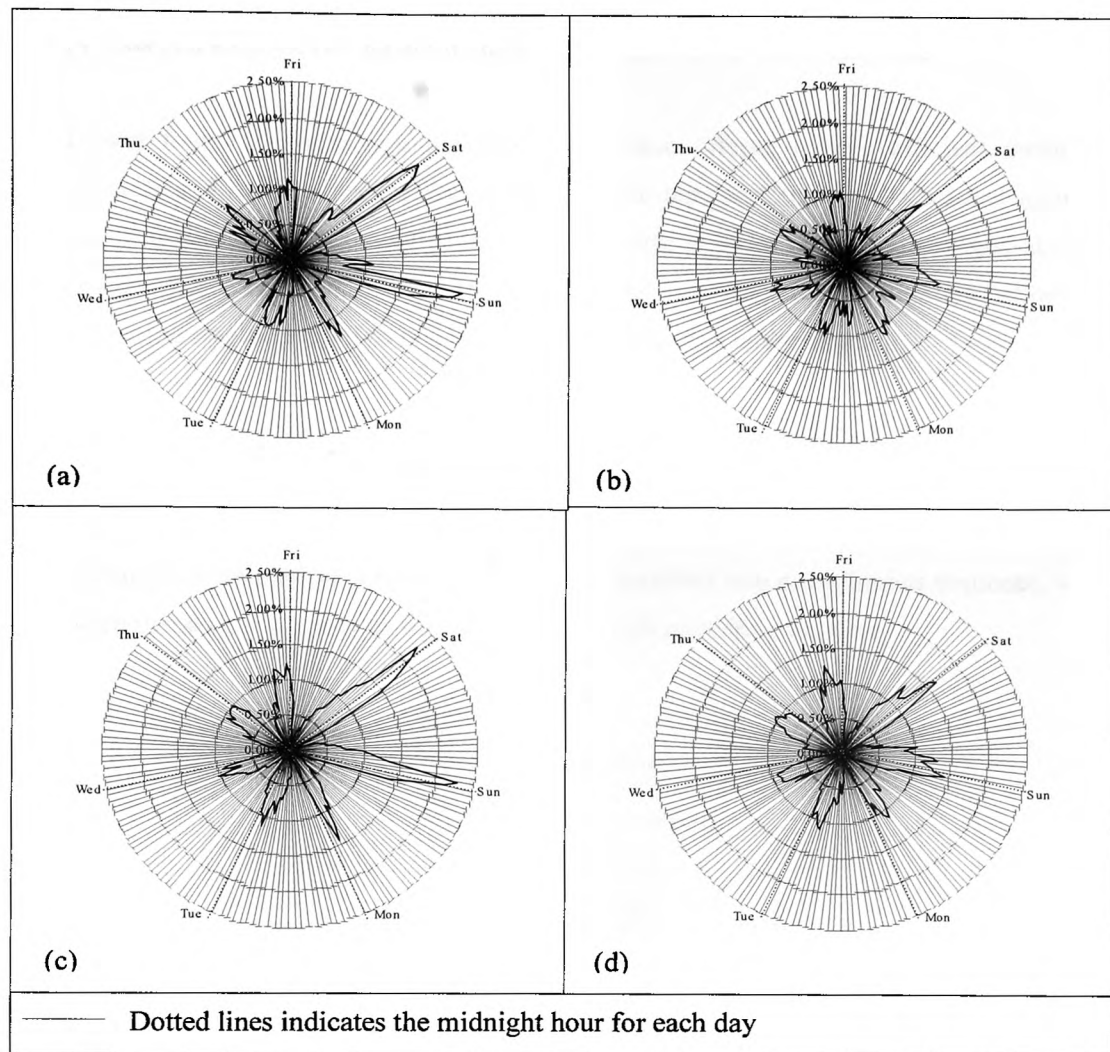


Figure 6-6 Percentage variation of incident by weekday and hour (a) c10 (b) c60 (c) d10 and (d) d90.

- **Spatio-Temporal**

Analysis of space and time in conjunction with one another was achieved using a variety of techniques to identify, and where feasible, quantify such interactions. From the previous two stages, a series of key findings were highlighted (for example, tendencies towards particular localities, times and commonalities across incident types); the aim of this stage was to attempt to unify the findings.

The first technique utilised the incident rate mapping (as utilised during the spatial analysis stage). For various granularities of time (month, week, day, hour) percentage changes for each region were calculated and appended with an appropriate colour

classification. The lifecycles of incident volumes per region could be analysed, while the periods and localities of apparent vulnerability and safety were identified.

It was found that although the above technique offered a time efficient method of generating visualisations for depiction of spatio-temporal trends, the manual nature of scrolling through each output of the time sequence reduced both its appeal and ability to communicate the patterns and trends exhibited. It was due to such limitations that animation was employed to improve overall quality and understanding.

The layer snapshot concept (Figure 6-7) uses a raster structure and has been applied to modelling the temporal dynamics of many forms of spatial data. Each snapshot is a representation of a particular time period and state of a particular phenomenon. The technique of animation combines the various snapshots into a continuous sequence, which when played can be used to simulate change of phenomena over time.

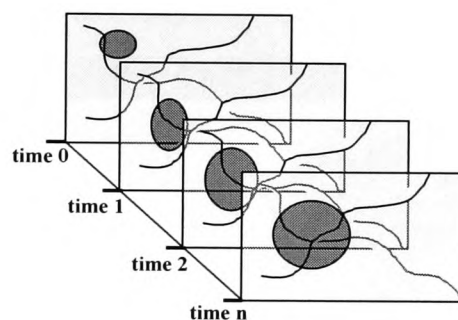


Figure 6-7 The snapshot model

Despite the various criticisms of the snapshot model concerning data redundancy (due to data replication) its use has proved a promising tool, providing basic spatial analysis and allowing fundamental questions to be addressed, such as:

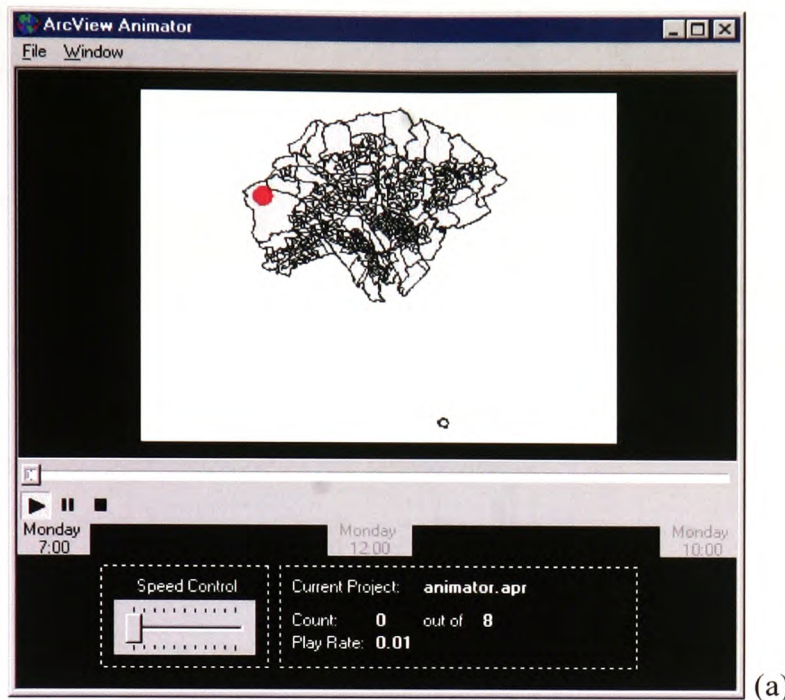
*What is going on?*

*Is there anything interesting?*

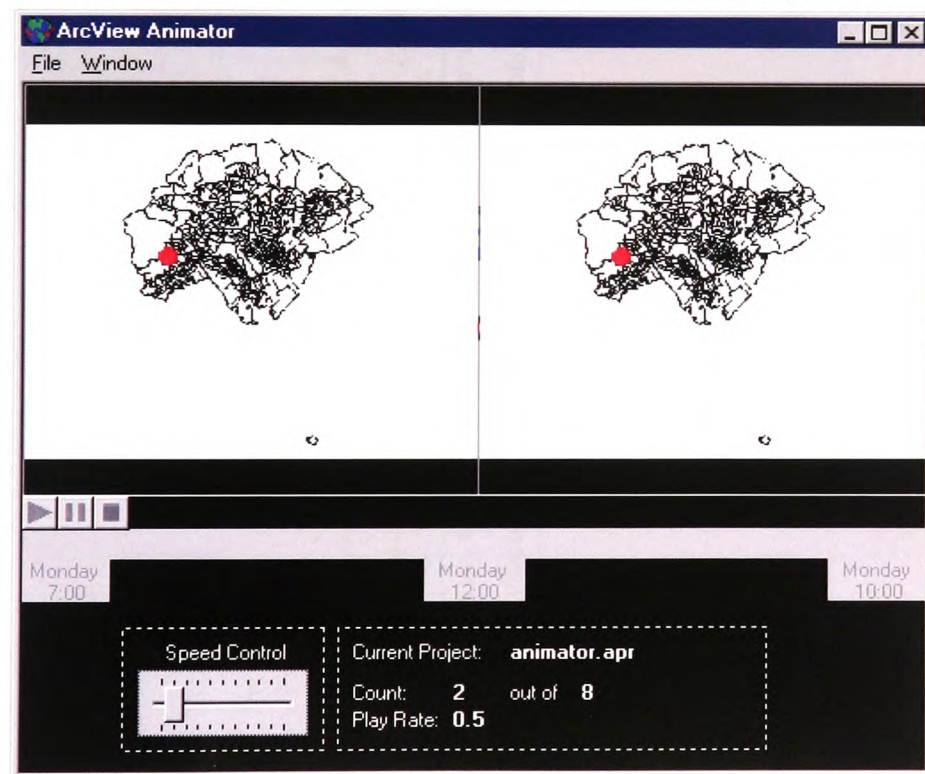
*What is happening where and when?*

*Are there any discernible patterns?* (Openshaw, *et al.* 1994: 132)

The very ability of this technique to address these most fundamental of questions must in turn highlight its validity in terms of a valuable, if not invaluable, exploratory procedure. The use and application is supported by various studies (Dorling and Openshaw 1991;



(a)



(b)

Figure 6-8 (a) Single view port (b) Dual view port animation player

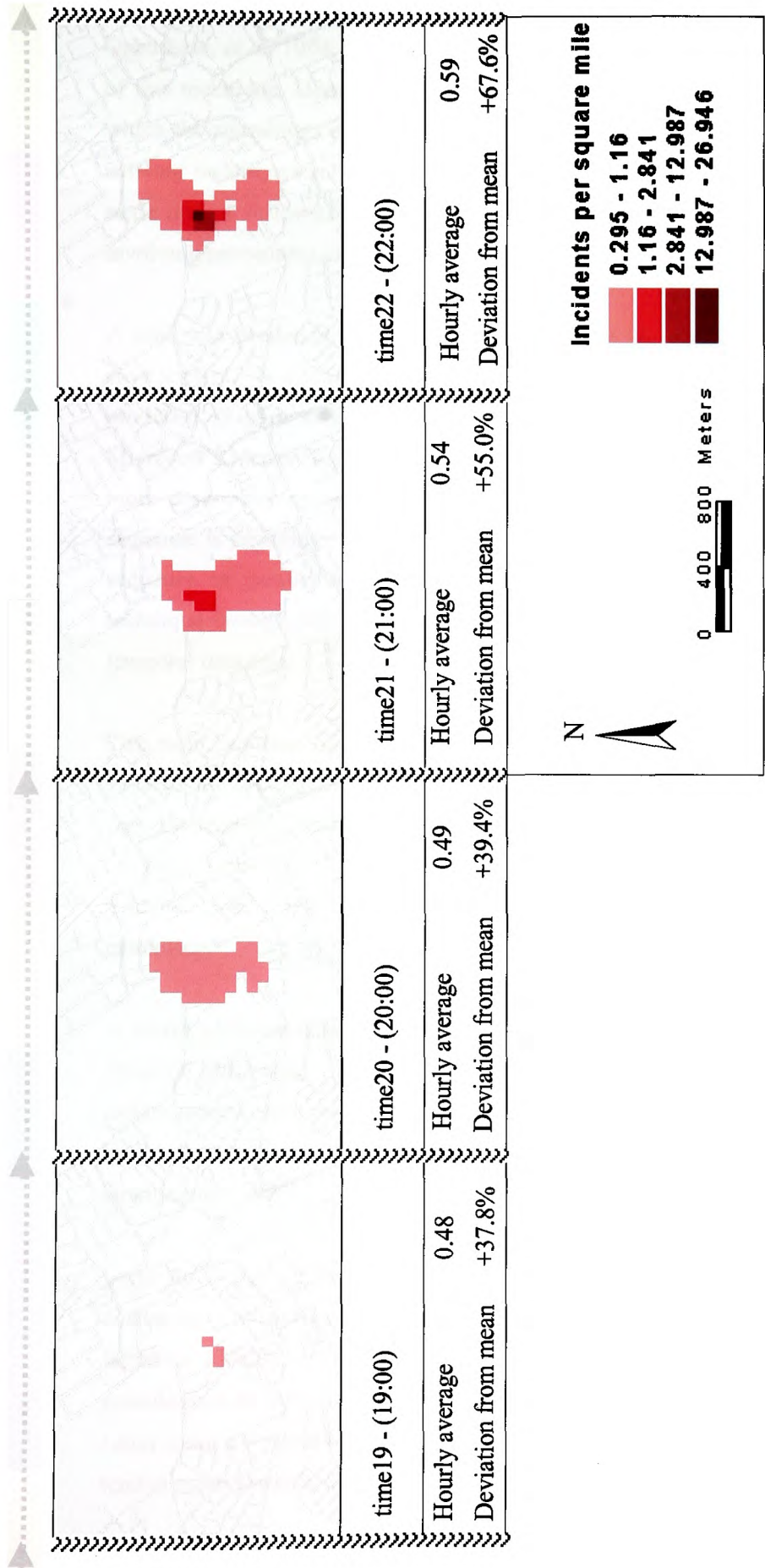


Figure 6-9 Subset from a typical animated sequence (specifically c10 focusing on the city centre region)



Openshaw, *et al.* 1994; Batty M. 1995). One of the main obstacles confronting wider use of this technique, alongside more traditional GIS techniques, is an efficient method by which the animations can be produced from within the GIS environment. On the whole, existing techniques rely on a lengthy, largely manual generation and conversion of the static images followed by frame-by-frame input in order to produce the animated output involving procedures external to the GIS environment.

A tool was produced to automate generation of the animated output within the GIS environment. In addition a customised facility for viewing the animated sequences was produced (Figure 6-8) as there is a tendency to utilise standard players such as Microsoft's Media Player. Here greater value was sought through utilising a player that is more descriptive of the material, such as including the time parameters for which the sequence is depicting. In addition the facility was designed to control more than a single sequence in parallel with one another, extending the exploratory capabilities of this technique through offering a tool to identify relationships between several distinct spatio-temporal data sets.

The main outcome of the animation investigation was the identification of distinctive patterns for each incident type (i.e. c10, c60, d10 and d90), and a series of hypotheses were developed. Figure 6-9 shows a typical subset from an animated sequence detailing the time, mean, and deviation from the mean for that particular hour. At this point, all evidence was interpreted, thus the next stage targeted the quantification of such patterning.

A series of quantitative tests were utilised to further interrogate the identified spatio-temporal patterning. Through implementation of such techniques the intention was to isolate typical movements or flows of criminal activity, investigating the concepts posed by Routine Activities and Point Pattern theory (Cohen and Felson 1979; Brantingham and Brantingham 1993).

Using the Moran's I statistic, the first stage was to quantify global evidence (city wide) of clustering or dispersion as depicted by the animation outputs. Results from the analysis failed to identify any marked changes in clustering or dispersion over the different granularities of time (i.e. day and hour). On this basis a distance analysis approach was taken using a Nearest Neighbour Analysis (NNA) to assess whether the point distribution tended more towards a clustered or dispersed arrangement than one might expect on the

basis of chance (Figure 6-10 Figure 6-11). In particular hour of day demonstrated the greatest variation in clustering and dispersion as calculated by the NNA.

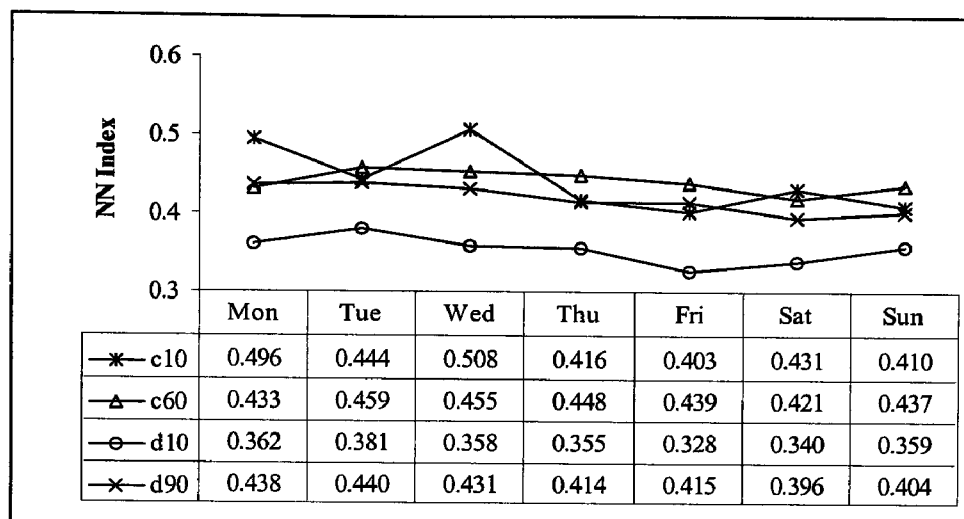


Figure 6-10 NNA by day (a) c10 (b) c60 (c) d10 and (d) d90

Each output demonstrated relatively high degrees of concentration throughout each time sequence. However, key epochs can be identified as exhibiting comparatively elevated levels of clustering. Such days (and in particular hour of day) correlated closely with evidence depicted through tabular and animation output, noting the tendency toward the weekend and the midnight hours.

To reinforce the global evidence of clustering and dispersion provided by the NNA, the second technique used a kernel density technique to identify evidence of local clustering. Borrowing from the field of ecology, certain techniques offer new opportunities when applied to the analysis and comprehension of criminal activity. One such technique is the delineation of the home range utilisation distribution (Worton 1987; 1989), where the objective is to assimilate a probability surface assessing the likelihood of an animal occurring at a given location given a series of sample points. It is possible to reapply these fundamental concepts to criminal activity, with the home range describing the region within which the criminals conduct their activities, and the utilisation distribution describing the vulnerability of particular localities to criminal acts.

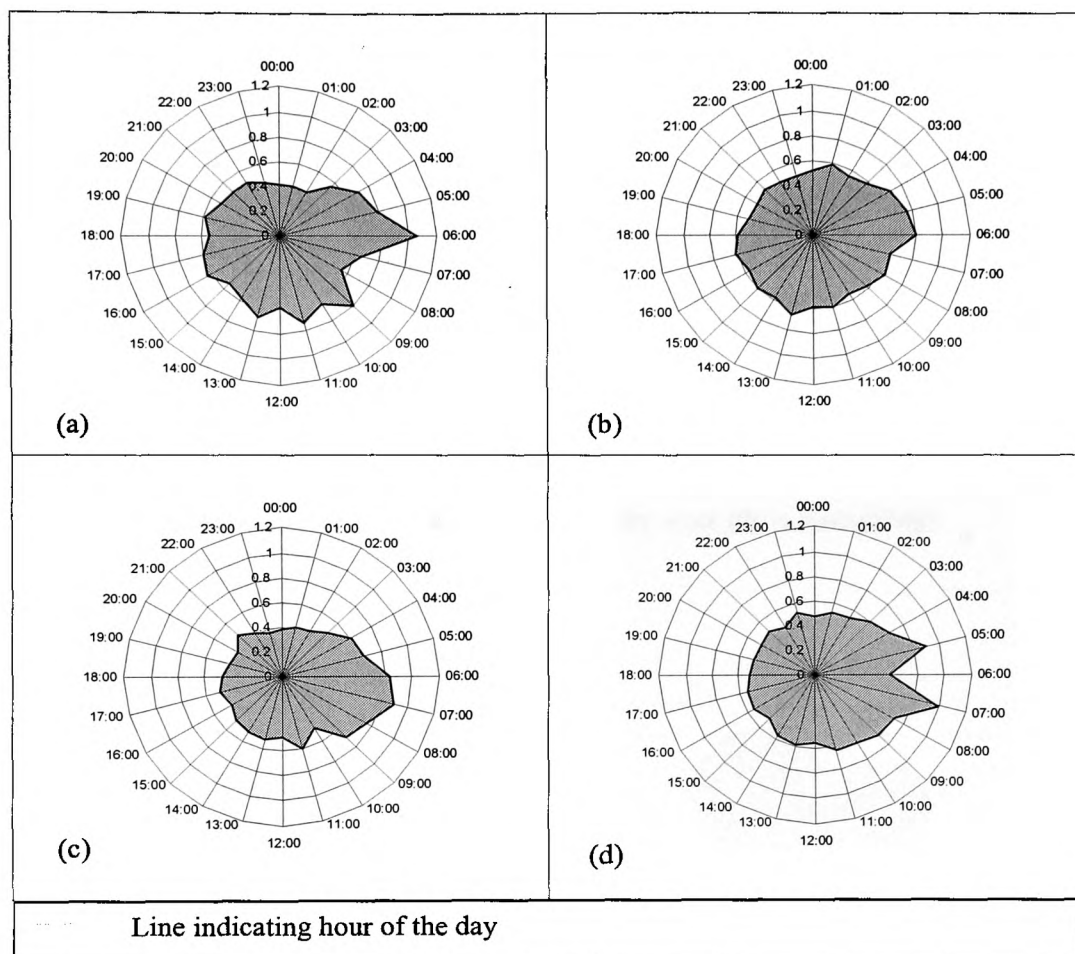


Figure 6-11 NNA by hour (a) c10 (b) c60 (c) d10 and (d) d90

In ecological studies the home range is determined as an area normally frequented by an individual animal. In the case of incident analysis, this is modified to define areas frequented by criminals committing crimes of the same type. Such areas can thus be considered as containing the essential elements (for example, suitable targets associated with minimal risks) supporting incident of a specific type, which is similar to the existence of sustenance and cover in the case of ecological studies.

Analysing both the day of week and hour of day, probability contours (50%, 75% and 90%) were calculated. The area that each probability covered was calculated and its percentage in relation to the total area of Cardiff determined. Graphs were then created that plotted the percentage against the temporal variable for each incident type (Figure 6-12 and Figure 6-13).



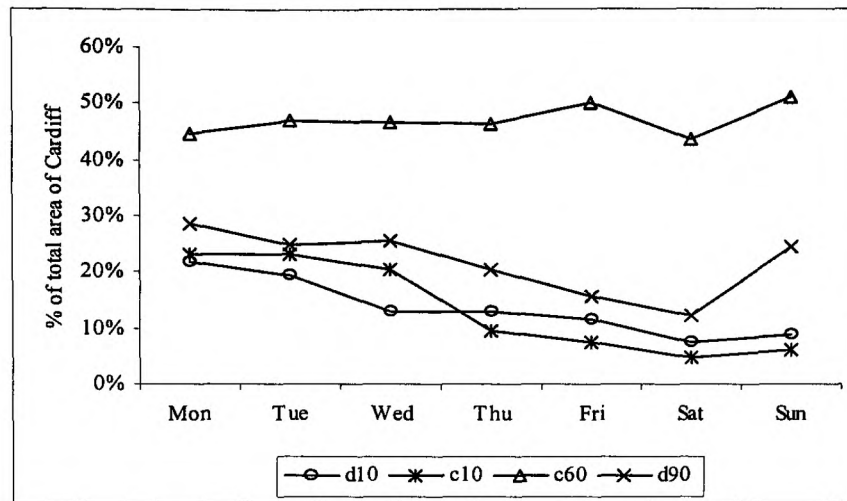


Figure 6-12 Variation in the utilisation distribution by week (90% probability)

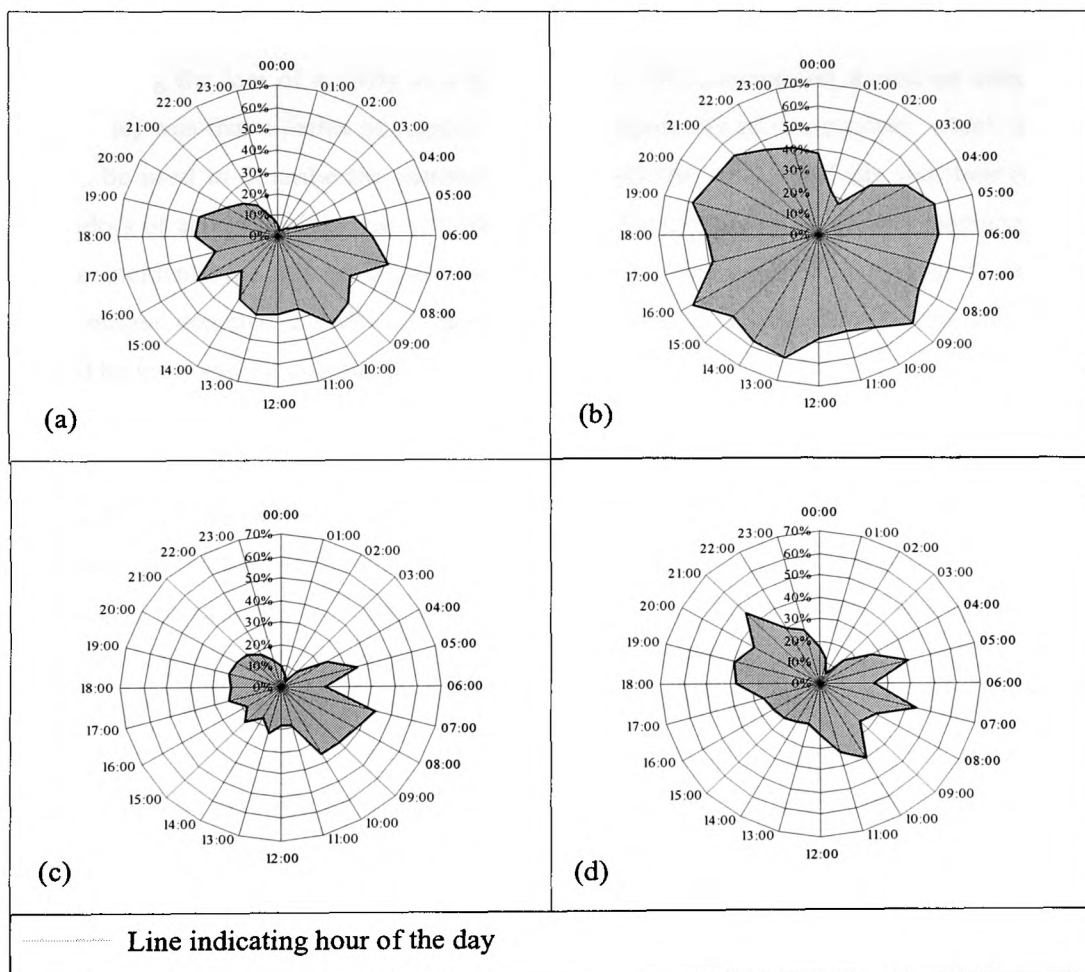


Figure 6-13 Variation in the utilisation distribution by hour (90% probability) for (a) c10 (b) c60 (c) d10 and (d) d90

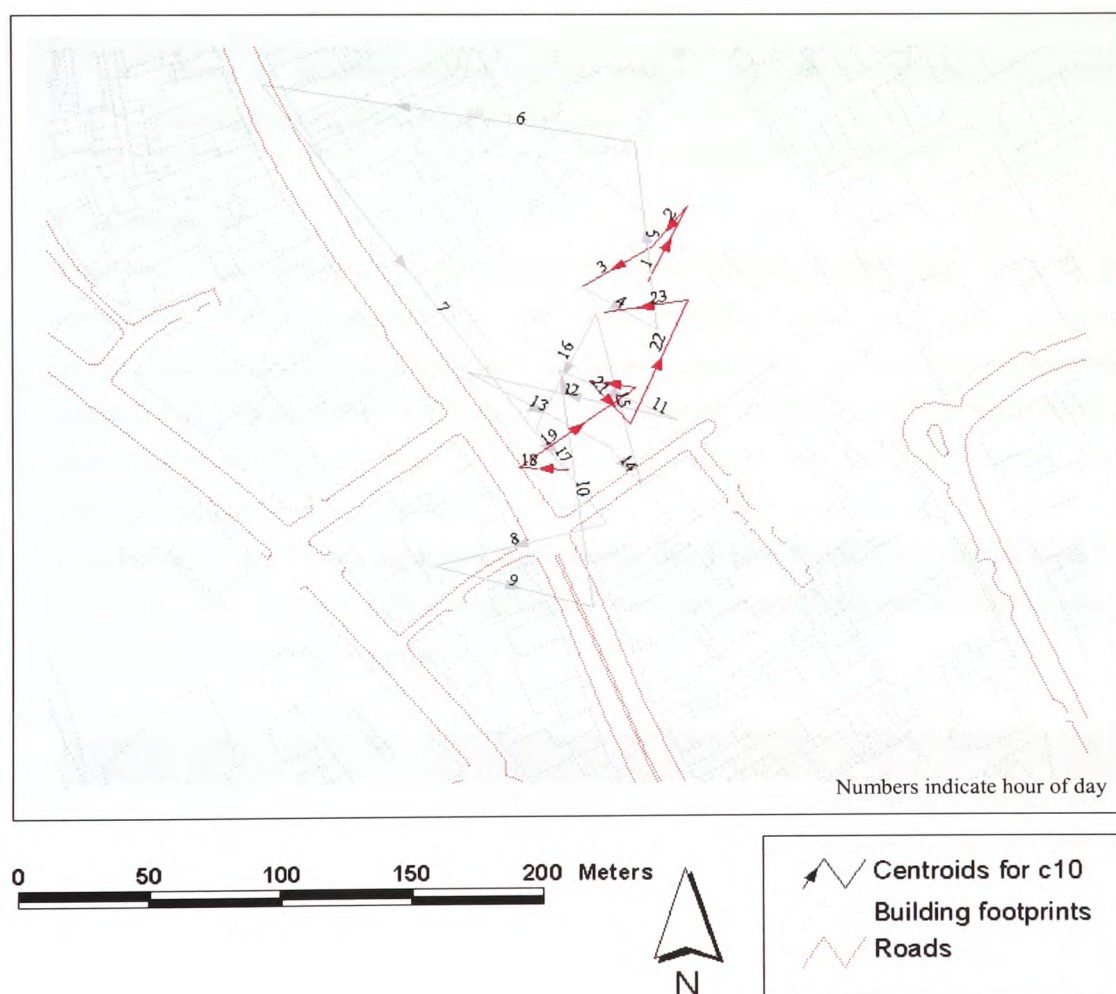
Results from the Moran's I and NN analysis in conjunction with those from the kernel analysis supported the animation evidence of a convergence of incidents (i.e. c10, c60, d10 and d90) in time and space and that is most marked during the weekend. The c10 incident type offered the most well defined patterning, where both temporal and spatial variables exhibited relatively strong clustering towards the midnight hours. This was especially the case over weekend days, during which time it was confined to relatively small areas of the study region. Results suggested that during the week, approximately 19% of the study region equates to the 90% probability area, decreasing to approximately 6% over the weekend. Similarly, between 23:00 and 03:00 the 90% probability area was approximately 4%, increasing to around 36% between the hours of 09:00 and 17:00.

The final type of analysis aimed to quantify the flows of incidents that were evident in the animations. To achieve this it was first necessary to isolate key areas of activity constituting the loci of activity in a given region. The assumption (based on animation evidence) was that a series of regions acted independently of one another, which in turn could be used to describe the common routine activities of individuals and their typical corridors of activity. This was achieved through the overlay of all density mapping for a given incident type and temporal granularity for delineation of hotspot activity extents. The output depicted a series of distinct regions, within which distributions and flows could be individually examined.

For an identified hot spot region and incident type, the mean centre was calculated and plotted. This was repeated for each time facet (days and hours) in the temporal sequence, the results focusing on the identification of centroid movements. Concentrating on the individual incident types, hotspot regions were identified for local flow modelling. Using the methodology described above, these were independently analysed, the results shown in Figure 6-14.

The city centre (as defined by the hotspot delineation process), comprised nearly one third of all violence against the person (31.7%), in addition to the highest concentration of leisure activity opportunities (shops, cinemas, restaurants, bars and night clubs and a major sporting stadium) across the city. For the other incident types the city centre constituted a smaller representation in their city wide distribution (d10 – 22.2%, d90 – 23.6% and c60 – 13.0%). This explains the extent to which this region witnesses large influxes and departures of individuals, both from city residents and those commuting

from surrounding areas. And it is these movements and concentrations of individuals resulting in a criminal act that can be quantified using the above technique. Their explanation was relatively simple, in that these flows are a reflection of an individual's routine or activity corridor through which they pass. The nature of urban morphology (for example, roadways and restricted entry to parks and shopping centres after certain times) constrain this activity, thus people are in effect funnelled along pre-defined corridors, whether this was a route to and from their residence or a transport hub. In addition, timing of events appeared to be strongly influenced by opening and closing hours of the various leisure activity nodes.



**Figure 6-14 Flows of c10 incidents within the city centre activity region.**

Animation and the flow analysis technique offer a way by which the hotspotting process (an already popular and established technique amongst many police forces in the UK, see Chapter Three) can be extended to advance explanations of spatio-temporal dynamics.

Ultimately these techniques can be applied to resource allocation decisions and to advise situational crime prevention programmes.

Key Findings	
•	Incidents (c10, c60, d10 and d90) and in particular c10 tends to be concentrated within small well-defined regions that both modify in area and distribution over various granularities of time.
•	As the volume of incidents increases the areas in which these incidents are committed does not proportionally increase i.e. the same areas are simply subject to more crime. This is particularly true of the c10 incident type.
•	Flows of incidents within the city centre appear to be strongly influenced by opening hours of leisure activity nodes

- **Contextual**

Time frequency plots were mainly used to identify *typical* distributions and trends in temperature, rainfall, sporting events, and school and public holidays, plotted against one another and the incident data for visual inspection. Employing a range of the hypotheses generated as a result of the temporal and spatio-temporal inspection, one process involved identification of extreme values (hours, days and months) and explanation through the various contextual data. One of the most prominent findings was the consistent correlation of major sporting events (specifically the Rugby World Cup) and increase in the volume of incidents. This relationship was also true of public holiday periods, specifically the New Year period.

Key Findings	
•	Public holiday periods, particularly New Year increased expected averages by up to 247% - disorder/disturbance; 115% violence against the person; 102% criminal damage (police) and 427% by A&E.
•	Major sporting events consistently increased expected daily averages by up to 125% (disorder/disturbance – police) and up to 289% in A&E records.
•	Weather offered a much less well-defined explanation, however appeared to provide contributory evidence for increases during bank holiday periods, in particular the August break with warmer drier conditions encouraging movements of individuals.

- **Discussion of police analysis**

Analysis of the police data focussed on four key stages (spatial, temporal, spatio-temporal and contextual), each of which was designed to assemble a profile of the study region in relation to each incident type. The four selected incident types each reflected similar spatial and temporal signatures where the city centre, weekends and the midnight hours played a major role in their explanation. Although no data was available, it was likely that a large proportion of these incidents were alcohol related, reflecting the negative element of the entertainment culture and its impact upon areas offering such facilities. In addition, opening hours appeared to play a major role, acting as an attractor on the basis of if the venue was either opened or closed raising issues of the currently imposed licensing laws. As the number of open venues reduces, individuals are drawn to a fewer number of locales. As a result, this tends to instigate movements of individuals between venues that typically culminate in movement to transport hubs (for example, train stations, taxi ranks and car parks) to return home. In essence, these movements place a variable demand on a changing resource level (i.e. number of open venues), which, during specific points in time (specifically weekends and midnight hours), may become overburdened with individuals wishing to gain entry to such venues. This is also true of the available transport options that reduce in variety towards the midnight hours placing greater demand on those remaining. Using this concept of resource demand helps to explain the temporal geographies of each incident.

The 2001/2002 BCS (Simmons 2002) was used for comparative purposes as is approximately coterminous with that of the police data coverage. However, the BCS does not provide a direct comparison to police incidents, but recorded crime (specifically wounding, common assault and vandalism) therefore was treated as a guideline only. The BCS corroborates some findings, indicating that violent incidents (descriptive of c10 incidents) tend toward the evening and night, with 43% occurring over weekends (37% police). The trend in violent crimes is one of decline, with the BCS indicating a decrease of 17% since 1999 compared to a value of approximately 8% decrease for the Cardiff data. Vandalism (criminal damage descriptive of c60) was reported to have risen by 11% whereas the Cardiff data suggested a 5% fall.

A limitation of the kernel technique in its calculation of area is that it fails to take into account the underlying geography issues of access and accessibility of regions. The technique does, however offer an insight into the spatio-temporal dynamics of criminal activity and can be used to quantify the global changing level of criminality.

#### 6.3.4 Comparison of Police and A&E data

Comparison of the two data sets affords an opportunity to investigate the approximate degree to which community violence is under reported by police statistics. Previous work has suggested that 95% of those sustaining injury through violence are treated in an A&E department (Shepherd, *et al.* 1993). Work by the TASC project (Targeting Alcohol-related Street Crime) based in Cardiff city centre and the Bay identified that assaults recorded by the A&E department constitute approximately 30% higher volumes than official assaults recorded by the police (Maguire and Nettleton 2002). The objective here was to establish the nature of any city-wide differences and whether these differences were consistent over various granularities of time and during specific events, for example, sporting internationals.

A cautionary approach was taken in this analysis, as the location of violence recorded by the A&E data (i.e. the catchment area), was unknown. In contrast, the police data was strictly defined by geographic boundaries (i.e. the city boundary). This element of uncertainty was reduced to some extent by previous work on violent crime in Bristol (Shepherd 1990), that suggested the majority of A&E assault victims are treated in the nearest available hospital. Therefore, the data should constitute an appropriate representation of violence occurring in neighbouring communities. An additional problem was that neither data set followed a common classification system. This fact, coupled with the geographic uncertainty, meant that the results presented were taken as only an approximate guideline as to the magnitude of difference between the two data sources. The classification issue is, however, subject to change with the introduction of an assault questionnaire (Goodwin and Shepherd 2000) through which it is planned that A&E departments will collect data comparable to Home Office categories of violence. This coupled with ethical approval, to include the locale of assault, will allow a more robust comparative analysis of the data to be conducted, and the shortfall of policing figures better appraised.

Analysis targeted the comparison of daily and monthly trends (Figure 6-15) and to some extent the volumes between assaults recorded by A&E to that of violence against the person as recorded by the police. A ten month period was used (maximum possible overlap given coverage of the data) where monthly and daily trends were found to exhibit very similar distributions, both indicating evidence of increased violence towards the weekend and both showing a distinctive rise during October 1999, with a likely explanation being the influx of people for the Rugby World Cup. Throughout the coverage period, A&E data consistently reported higher volumes of violence (40.9% difference), which are becoming more

pronounced during weekend periods (59.2% higher than police recorded violence). Both are greater than that reported by TASC. Inspection of the daily log revealed that in 73% (n=224) of all cases, A&E recorded a higher incident volume. Only 11% (n=32) were within a 10% margin of one another. Some of the highest differences were experienced during the rugby internationals (up to 250% variance) and New Year period (up to 226%). As previously mentioned, these disparities can potentially be used to provide an approximate indication of differential violence detection rates, but, however, cannot be used as a substantial and robust measure.

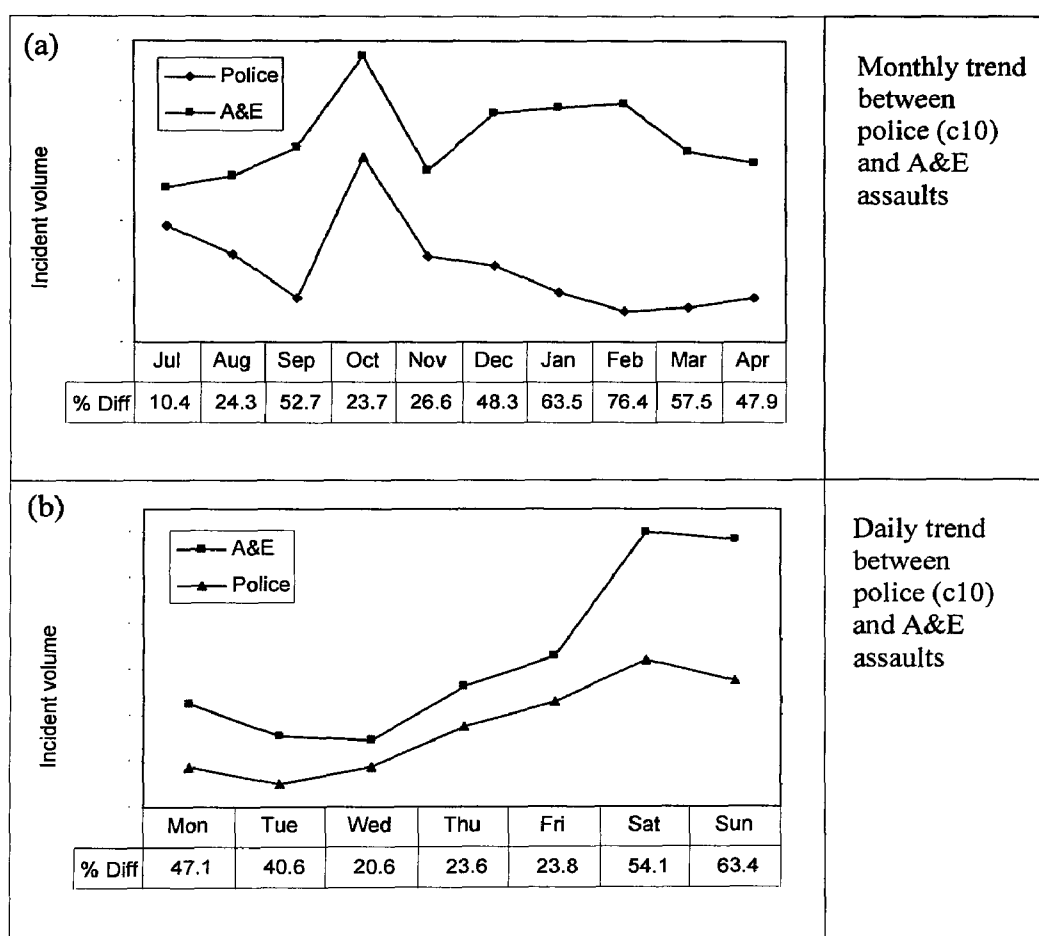


Figure 6-15 (a) Monthly trend (b) Daily trend

The duration of data used in this comparison provided some interesting insight. However, future work in this domain (ethical issues aside) should target investigations similar to those used in the Bristol study, incorporating locale and hourly timings of assault in addition to a socio-economic appraisal of assault victims. Combining such additional information with comparable data from police records offers the platform from which a much richer survey could be conducted. The study could follow a similar method to that advocated by the TASC

project, but encompassing the entire City and employing much greater use of GIS tools (i.e. hotspot animation and flow analysis).

#### **6.4 Summary and Conclusions**

Overall this Chapter has presented a series of investigative techniques for inspection of police incident and A&E injury data that were coupled with a variety of contextual sources to help broaden their explanation. The content and time span of each of the two data sets offered an opportunity to assess a variety of exploratory techniques to assess patterning and relationships. The A&E data set covering a five-year period combined with a range of additional descriptive variables (i.e. age and gender) was analysed using traditional statistical techniques investigating aspects of seasonality and associations with additional calendar and meteorological variables. The police data in contrast covered a two-year period offering spatial, temporal and incident type information. As a result, the analysis focused on the determination of temporal geographical patterning. In addition, a comparison was made between police and A&E, where the focus was the identification of similarities and differences in incident volumes.

- Analysis of A&E data has shown a distinct weekly pattern particularly for the 18-30 male category and one that appears to be becoming more pronounced. The female 18-30 age category also saw a marked rise, but males remained the predominant gender constituting 70% of all assaults.
- Analysis of police data highlighted similar tendencies towards the weekend days and midnight hours, but none of the incident types investigated indicated an increasing trend to that identified in the A&E data. However, comparison of police (c10) and A&E data highlighted relatively substantial differences in reporting, which helps to explain the overall trends each exhibited.

Spatio-temporal analysis using animation and quantitative techniques (i.e. Moran's I, NN, and kernel density methods) highlighted the limited locales that are subject to the majority of incidents and their modification over weekly and hourly cycles. The c10 incident type showed the most concentrated patterning (in both space and time), c60 in comparison affecting a much greater area and spread more evenly over time.

Analysis of calendar events and in particular sporting events and the New Year period was shown to impose a marked influence, increasing the number of both police and A&E



incidents above expected levels. Use of other explanatory variables (for example, meteorological parameters and other holiday periods) failed to provide a consistent explanation for the extent of data analysed here, but these could be used in an anecdotal context. Thus, for a specific point, a sophisticated explanation could be derived using a combination of explanatory seasonal and calendar variables. For example, the amalgamation of a hot summer, extended public holiday weekend, in addition to a scheduled sports event may provide a fuller explanation of incident and injury volumes.

Seasonality is a well-established explanation of crime level fluctuations over monthly and longer time aggregates. Exploration of the Cardiff data (A&E and police) failed to confirm any evidence of seasonality, with smaller aggregates of time (weekly and hourly) offering the most well defined explanation. This suggested that there was a city level model, whereby seasonality was over powered by the weekly cycles producing a more dominant and distinguishable effect on injury and incident<sup>1</sup> figures. This cycle appeared to be strongly influenced by entertainment that was confirmed by the locale, timings and association with specific calendar events. Analysis of other cities coupled with regional and national level data is now required to support this phenomenon.

---

<sup>1</sup> For those incidents investigated (i.e. c10, c60, d10 and d90)

## 7 MODELLING CRIME DATA FOR PREDICTION

*This Chapter presents a methodology whereby crime can be geographically predicted for areas that typically transcend traditional policing boundaries. A geographical crime incidence-scanning algorithm was developed to cluster locales with relatively high levels of crime (hot-spots,) providing sufficient data for training ANNs. The Gamma Test (GT) was applied to each cluster to assess suitability for predictive modelling. Following the results from the GT the outcomes of two forecasts using ANNs and linear regression are presented. The Chapter concludes by a discussion of the merits of the methodology along with its potential for future development.*

---

### 7.1 Introduction

One of the next steps in the enhancement of the crime mapping toolbox is the development of predictive facilities, where extrapolations from past and current crime patterns are used to formulate projective estimates of where and when crime might ensue. Olligschlaeger (1997) provides an overview of existing forecasting techniques, concluding that the time, level of user interaction and the expertise that each demand is unrealistic for implementation in an operational policing environment. In addition, the inherent inflexibility to dynamically adapt to change would compromise their viability in policing.

A system that intelligently interrogates a constantly updated database of crime incidence, providing indicators of where and when crime is likely to be highest, would be of great benefit in real-time police resource allocation.

Prediction requirements for the police have been classified into three main categories according to the period of time involved (Gorr *et al.* 2000): short-term (tactical deployment); medium-term (resource allocation); and, long-term (strategic planning). Prediction can help prevent crime in that it facilitates the optimal allocation of limited resources. However, Gorr, *et al.* (2000) have shown that the predictive power of forecasting models is a product of the incidence count utilised and that these are generally low in relation to crime type, time and space, and subject to randomness.

## 7.2 Initial attempts at modelling

Crime is a complex phenomenon, thus a certain level of generalisation is required in order to model its dynamics. Initial attempts used policing boundaries as a basis from which predictive models were constructed. For each region (shown in Figure 7.1), point level data was aggregated to form training sets from which individual ANNs were trained. However, only limited success was achieved, with only one region (the city centre) yielding enough data to permit accurate modelling. In addition, the 18 areas for which the predictions were made tended not to reflect the underlying crime distributions, thus limiting their applicability for resource allocation. Such results prompted the development of CLAP (Cluster Location Analysis Procedure), which moves away from the use of traditional boundary structures towards a dynamic boundary derivation methodology from which predictive models can be built. (Figure 7.1 illustrates police sector boundaries for Cardiff using the CLAP interface.)

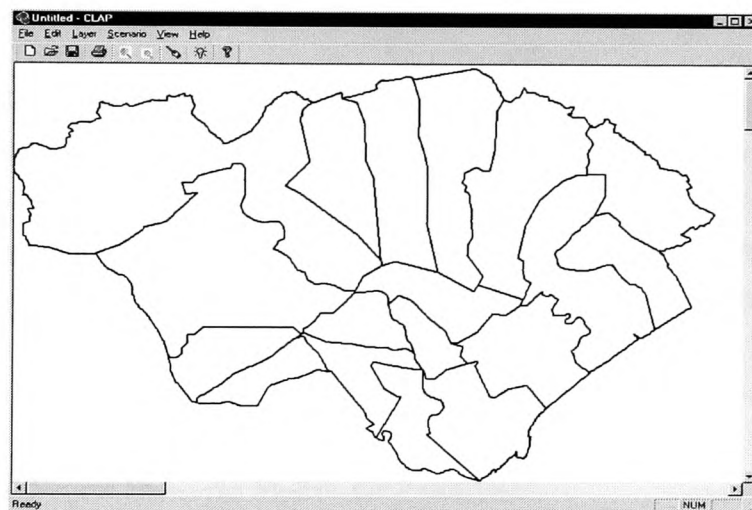


Figure 7.1 Policing boundaries (sector level) for the study area

The methodology developed follows three key stages (summarised in Figure 7.2). First, spatial analysis is used to define geographical clusters. Second, cluster modelling is used to determine data quality for each of the clusters. Third, the results of the previous stages are used to define predictive models.

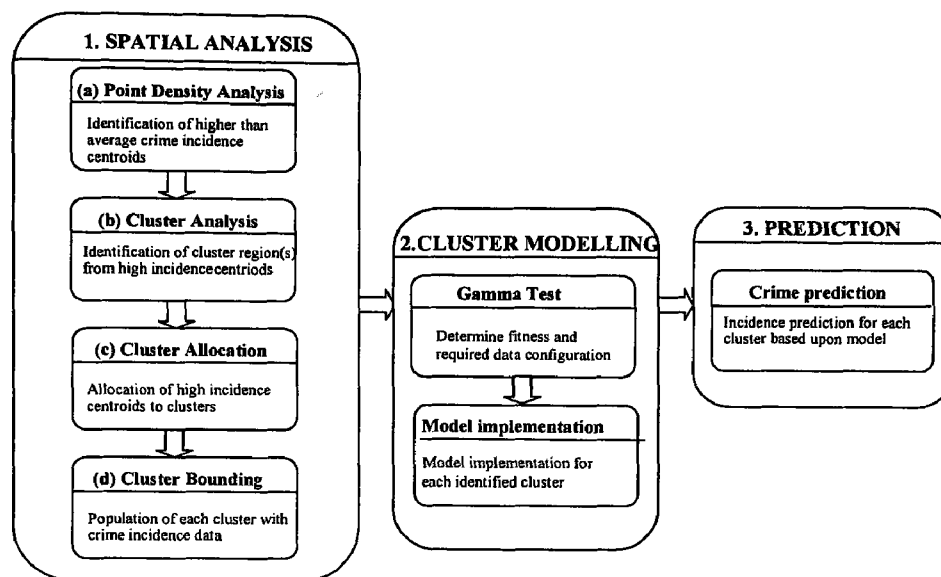


Figure 7.2 The CLAP/GT/Predictive process model

This Chapter details a potential prediction framework for short-term, tactical deployment of police resources. The objective is the identification of areas where the levels of crime are high enough to enable predictive models to be produced. This work differs from other recent studies dealing with hot-spotting methods (Ratcliffe and McCullagh 1999) and with their statistical significance (Chainey and Reid 2002). Whereas these papers employ hot-spotting techniques as a means of visualising and comprehending crime distributions, here their utility is extended to use identified hot-spot regions as the foundation for predictive models.

### 7.3 Artificial Neural Networks Models for Prediction

An ANN has been described as “*humanity’s attempt to mimic the way the brain does things in order to harness its versatility and ability to infer and intuit from incomplete or confusing information*” (Tazelaar 1989: 214). More specifically, they can be used to learn complex relationships for the recognition of patterns.

An ANN (Zurada 1992) is first “trained” to identify the relationship between a series of input vectors and their corresponding output vectors. Input vectors are repeatedly presented to the network one at a time, each element of the vector corresponding to a different neuron in the first layer of the network (Figure 7.3a). These inputs are then multiplied by a corresponding weight before being forwarded to every neuron in the next layer of the network. The forwarded values are multiplied by the weights associated with each neuron (Figure 7.3b) before being summed together with all the other values being sent simultaneously to that

neuron. The summed values (one per neuron in the layer) are then forwarded, usually via a transformation function, to the next layer of the network. The output from the final layer corresponds to the network's calculation as to what the output vector should be. Initially all the weights in the network are set to random values and the network "learns" by adjusting the weights in such a way as to reduce the difference between the network's calculation of what the output vector should be and the actual value.

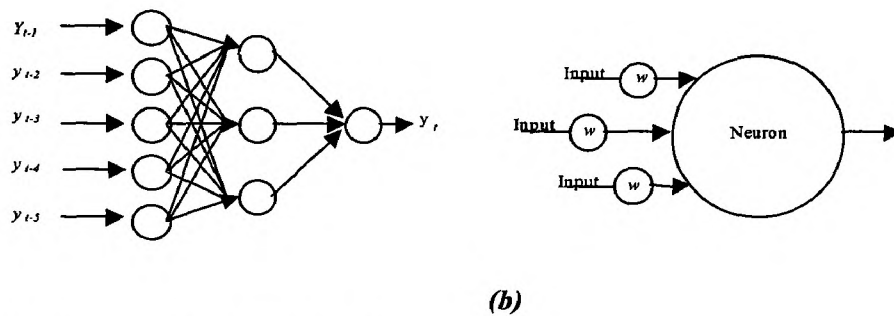


Figure 7.3 (a) A diagram of an ANN (b) A diagram of a neuron.

In the case of crime level forecasting, the input vectors and output vectors are counts of crime. The network consists of multiple inputs  $y_{t-1} \dots y_{t-n}$  and a single output  $y_t$  (Figure 7.4). The time series is arranged into a series of windows each constituting a portion of the entire data set. A window (of a specified length) is passed over the data and the ANN used to extract salient features of the series.

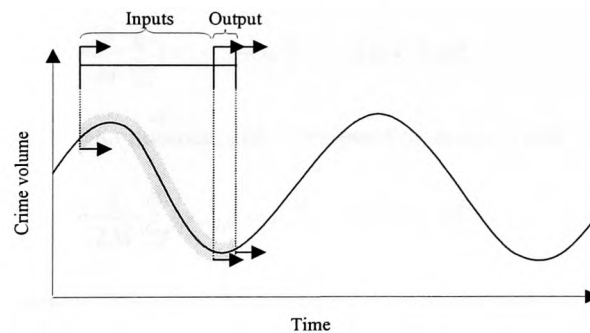


Figure 7.4 Moving window technique

Network testing involves presenting it with a series of input vectors for which the tester, but not the network, knows the corresponding crime levels. The answers given by the network as to what it determines to be the level of crime, given the presented input vector, can then be used by the tester to determine the robustness of the training process. If the robustness of the network is deemed sufficient, the network can be used in a truly predictive capacity. Here the

network is presented with an input vector for which the output is not known and its answer is assumed reliable. If during testing, the robustness of the ANN is found to be below acceptable limits then the topology of the network (that is the arrangement and number of nodes) and the coding of the input vectors needs to be re-examined.

#### 7.4 The Gamma Test

The Gamma (or near neighbour) test (Stefánsson 1997) is a non-parametric technique that allows an estimation of a model complexity and quantification of its inherent noise. The test estimates the best Mean Square Error (MSE) that can be reached when modelling the data using any smoothing method, such as an ANN. The technique can be summarised by:

$$y = f(x) + \varepsilon \quad (1)$$

where  $y$  is the given output of an unknown function  $f$  and  $\varepsilon$  is that part of the output not account for by  $f$ . Thus, where the function  $f$  is unknown and  $\varepsilon$  is statistical noise, the GT estimates the variance of  $\varepsilon$ . The GT is applied to all  $M$  data points  $(x, y)$ , based upon distances between each  $i^{th}$  data point and their nearest  $p$  neighbours. Specifically, the GT is derived from the input vector:

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N[i,k]} - x_i|^2 \quad (1 \leq k \leq p) \quad (2)$$

where  $|\dots|$  denotes Euclidean distance, and corresponding output value:

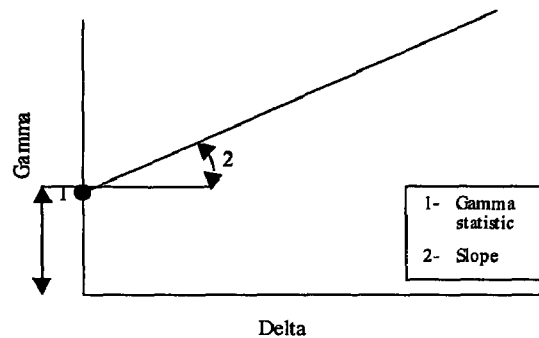
$$\gamma_M(k) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i,k]} - y_i)^2 \quad (1 \leq k \leq p) \quad (3)$$

which are found and fitted with the regression line:

$$\gamma = A\delta + \Gamma \quad (4)$$

of the points  $(\delta_M(k), \gamma_M(k))$  ( $1 \leq k \leq p$ ), (Durrant 2001). Here the  $p$  nearest neighbours, used by the GT procedure is fixed and bounded ( $2 \leq p \leq M$ ).

The resulting graphical output (specifically the regression line, shown in Figure 7.5) provides two key indicators. Firstly, the vertical intercept  $\Gamma$ , returned as the estimate for the variance of  $r$  the intercept on the  $y$  or Gamma axis, offers an estimate of the best achievable MSE though use of any smoothing technique. The second, the gradient  $A$ , offers an indication of model complexity (where a steeper gradient indicates a model of greater complexity). Results can indicate variations in the two variables (for example, estimates of low MSE being associated with a high level of complexity), with the preferred scenario being a low MSE and shallow gradient.



**Figure 7.5 The Gamma Statistic and the Gradient/slope**

Using the estimated MSE, a useful test is to establish the minimum quantity of data points required to model the function. This is accomplished by conducting the GT on an increasing sample size ( $M_1 \dots M_m$ ), (known as the M-Test) and plotting the Gamma value against that of  $M$  (see Figure 7.14 and its explanation later for a fuller description). In an ideal model the output may exhibit large variation in Gamma at small values of  $M$  but ultimately stabilising at a higher value of  $M$ , which is indicative of the true noise variance inherent within the data. The region of asymptotic values of Gamma against  $M$  identifies the minimum data required to establish best possible accuracy in prediction.

The two measurable outputs from the GT offer a basis from which an ANN model can be assembled and trained, given that an estimate of the best possible MSE has been provided by the GT. Thus, the model can be trained to the point where the estimated MSE is reached; and so over-fitting can be avoided. (For a more comprehensive discussion on the GT see Jones, *et al.* 2002.)

### 7.5 Empirical evidence and output

For this analysis a similar technique to the GAM/1 geographical analysis machine developed by (Openshaw *et al.* 1987) was used, augmented to allow the clustering of centroids of high incidence. This four-stage process consists of:

- point density analysis;
- geographic representation and cluster analysis;
- allocation of centroids to clusters;
- relation of incidents to cluster boundaries.

The same police incidents as were analysed in the previous Chapter, (violence against the person, criminal damage and disorder/disturbance/ nuisance) were used as the foundation for

predictive models present in this Chapter. The exploratory findings provided a baseline for validation of ANN output.

### 7.5.1 Stage one, point density analysis

During this initial stage, an analysis of the crime data, based on the hypothesis that areas with crime incidence greater than the average are significant, is carried out. The algorithm that looks for areas of greater than average crime incidence is shown in Figure 7.6 and the ensuing results for the test data set are illustrated in Figure 7.7.

```

ScanRadius is the radius of circle within which points will be counted.
StepSize is the distance that the centroid will be moved at each iteration.
StartX is the bottom left longitudinal start co-ordinate.
StartY is the bottom left latitudinal start co-ordinate.
EndX is the top right longitudinal end co-ordinate.
EndY is the top right latitudinal end co-ordinate.
Average is the number of crimes per area of circle.
For X = StartX to EndX step StepSize (moves longitudinally across the map)
  For Y = StartY to EndY step StepSize (moves latitudinally up the map)
    Count the number of points within ScanRadius
    If Count is greater than Average then
      Add co-ordinates to centroid list.
    Endif
  Next Y
Next X

```

Figure 7.6 Incidence density algorithm

At each iteration, Count is calculated by iterating through the crime file, incrementing a counter each time a crime is found to be within *ScanRadius* of the current centroid (*X*, *Y* co-ordinate), making use of Pythagoras:

$$\text{if}(\text{ScanRadius} > \sqrt{(\text{CrimeX} - X)^2 + (\text{CrimeY} - Y)^2}) \quad (5)$$

then add 1 to *count*

where *CrimeX* and *CrimeY* are projected co-ordinates of the crime incident.



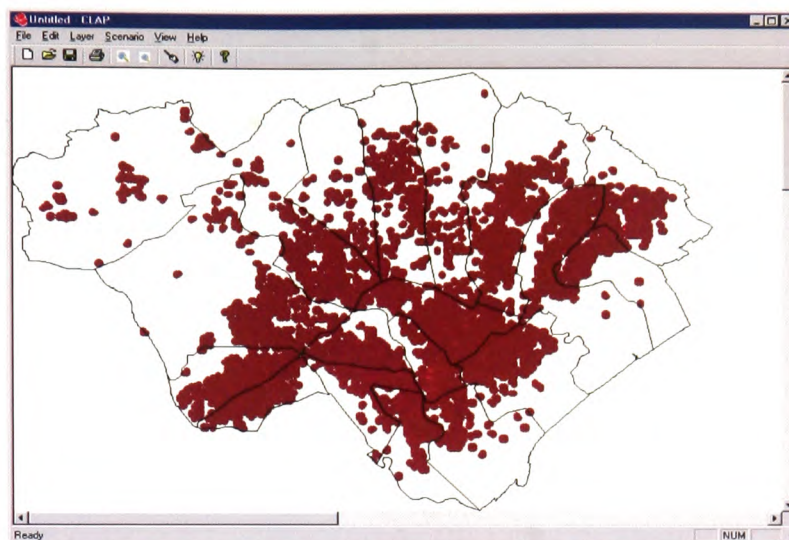


Figure 7.7 Centroids of higher than average crime incidence

### 7.5.2 Stage Two: Geographic representation and cluster analysis

At this stage, a heuristic approach is taken to determine the level of crime incidence required for a cluster to be considered salient. The heuristic rules utilised to make this determination are based on an assumption that most incidence of crime tend to be concentrated within relatively small geographic areas. This assumption is supported by exploratory evidence presented in the previous Chapter.

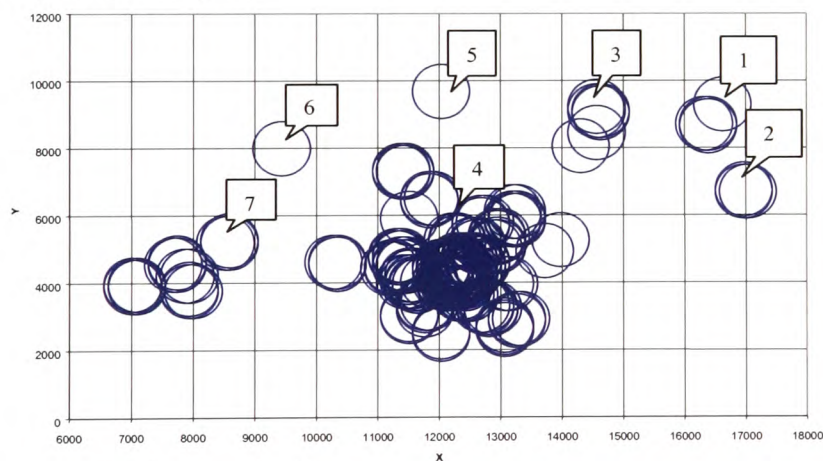


Figure 7.8 Clusters of high crime incidence

Given this, a scatter graph representation of the geographical data was utilised to heuristically increase both the density of centroids displayed and the radius of the area associated with that centroid. Therefore, as the density of the centroids displayed increases, the radius of influence associated with that centroid also increases. Experimentation resulted in the radius of influence, or *gravity*, being set to  $density \times 20$ , where *density* is the count of crimes associated with the centroid during stage one of the analysis process.

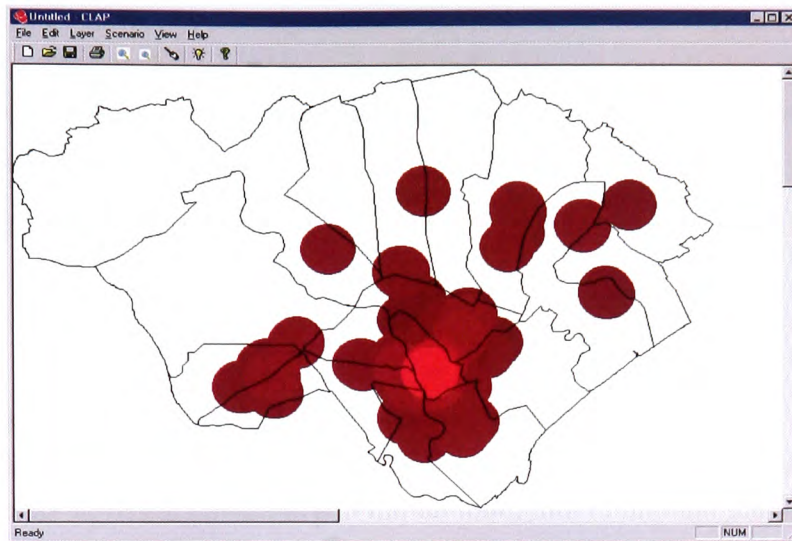


Figure 7.9 Clusters of high crime incidence, displayed using the CLAP interface

User interaction resulted in a centroid density of 40 being selected, resulting in the identification of seven clusters of interest (illustrated in Figure 7.8 and Figure 7.9). A list containing each of the salient centroids is then produced.

### 7.5.3 Stage Three: Allocation of centroids to clusters

During this stage, those centroids that should be grouped together to form clusters are identified. The density and gravity parameters and the centroid list generated in stage two forms the basis for this iterative procedure, outlined in Figure 7.10:

```
Enumerate each centroid and define each as being a potential cluster
```

```
While a cluster is growing
```

```
    For each cluster
```

```
        Attempt to Expand the cluster.
```

```
    Next Cluster
```

```
Wend
```

```
Remove duplicate clusters from list.
```

Figure 7.10 Centroid clustering algorithm.

*Expand* is a function that takes each cluster in turn (the list of centroids) and checks every other centroid to assess if there is an *Intersection* with a member of the cluster. If a centroid is found to intersect with a member of the cluster not already a member of the cluster, then it is added to the list.

*Intersects* is a function that takes two centroids and a distance value (*gravity* defined in stage two), and returns true if the distance between their centres is less than sum of the radii, according to the following equation:

$$\text{if} \left( (gravity * 2) > \sqrt{(c2.x - c1.x)^2 + (c2.y - c1.y)^2} \right) \text{ then} \quad (6)$$

*true else false*

where  $c_n.x$  and  $c_n.y$  are the central co-ordinates of each of the centroids.

#### 7.5.4 Stage Four: Relating incidents to cluster boundaries

Finally, each of the clusters is populated with data ready for training a series of ANNs (one per cluster). The algorithm utilised to perform the population is outlined below:

```

For each cluster
  For each centroid within the cluster
    For each crime
      Add the crime to cluster and add one to the count associated with this cluster if
      crime occurred within the centroid and the crime has not already been added.
    Next Crime
  Next Centroid
  Add Output results to unique cluster list
Next Cluster

```

Figure 7.11 Incident to cluster boundary allocation algorithm.

Equation (6) is again used to determine if a crime falls within the radius determined by the *gravity* value (800m). Each crime record contains a unique identifier, the cluster it belongs to and the weekday during which the crime was committed. In addition, each cluster record has a unique identifier, a list of its member centroids and a total crime count (shown in Table 7-1).

Cluster	Centroids	Crime Count
1	0, 4, 5, 6	1254
2	1, 2, 3	50
3	7, 8, 9, 10, 11, 12	954
4	13-109, 111-148, 165, 166	4097
5	110	161
6	149	228
7	150-164	2094
	Total Clustered:	8838
	Total Violent Crime:	18498

**Table 7-1 Crime incidence by cluster**

### 7.6 Application of Gamma test to the cluster data

Taking the results from the cluster analysis, two techniques were used to model the clustered data. The first sought to model day of week against crime volume, the second treated the data as a continuous times series using a windowing technique (Figure 7.4). Initial attempts using the first technique failed to achieve the accuracy accomplished by the second. The windowing technique was selected as the preferred procedure on the basis of the short-term forecasting problem.

### 7.7 Forecasting using Artificial Neural Networks

Implementation of an ANN model requires careful consideration of a series of model parameters each potentially impacting upon model stability and efficiency. These included decisions that concern architecture type (number of input/output nodes and hidden layers), selection of training algorithm and volume of data to be used for training and testing.

The ANN used comprises of an input layer (corresponding to the length of the input vector), an output layer, providing the forecast value, and two layers of hidden nodes. Previous research has indicated that use of a single hidden layer is sufficient to learn any complex non-linear function (Hornik 1991). More recent work, however, suggests that two hidden layers can produce a more efficient architecture than is possible using a single hidden layer (Srinivasan *et al.* 1994; Zhang 1994).

### 7.7.1 Network topology

Modelling the time series as depicted in Figure 7.4 involved passing a moving window of a specified number of inputs (initially set to a relatively large number to facilitate the increasing embedding test explained later) and a single output across the entire data set. The length of the window varied according to each cluster. The objective here was to optimise the number of input values required to make a reliable prediction.

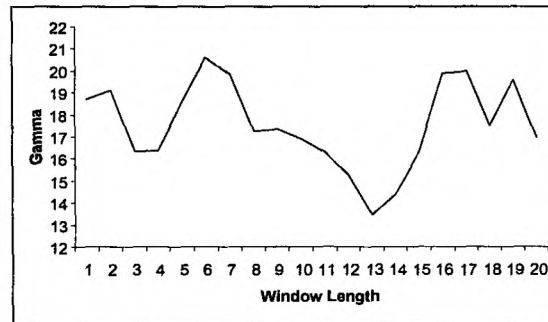


Figure 7.12 Example of increasing embedding (City Centre, cluster Four)

Window length was established by conducting an embedding test that calculates the Gamma statistic for an increasing number of data points moving along the entire time series. The number of inputs with the lowest Gamma value that produced the Gamma statistic closest to zero (for example, 13 for the city centre cluster shown in Figure 7.12) can then be found.

### 7.7.2 The number of nodes in the hidden layers

The optimal number of nodes in the hidden layer was found empirically.

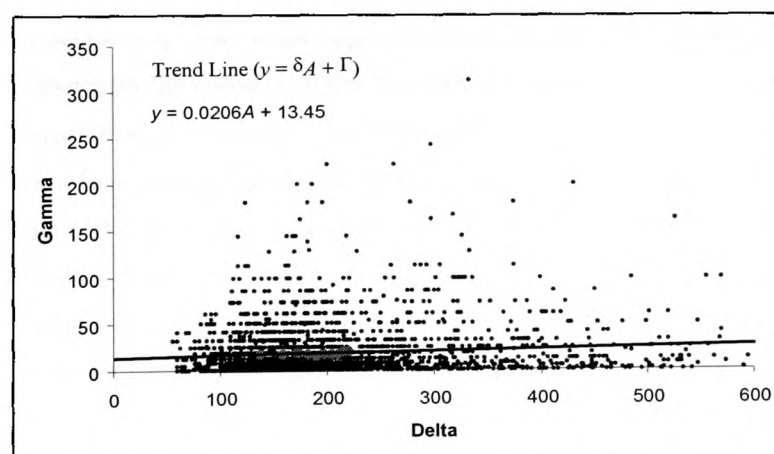


Figure 7.13 Gamma Scatter Plot showing the trend line

Initial large numbers of nodes ( $2N+1$  split evenly between both hidden layers, where  $N$  is the number of inputs) in the hidden layers were incrementally reduced to a minimum whilst maintaining acceptable forecasting capabilities. The shallow gradient (shown in Figure 7.13)

suggested that a relatively few number of hidden nodes in proportion to the  $2N+1$  rule would be sufficient to model the underlying function and this proved to be the case.

A single output node was used that is representative of the one step ahead forecasting horizon.

### 7.7.3 Terminating the training procedure

As over-fitting is a widely accepted problem associated with utilising ANNs, the GT's ability to accurately measure the 'noise' within a data-set and, consequently, the point at which training should stop, provides a significant utility for practitioners. Over-fitting occurs because the ANN will eventually attempt to fit all data encountered, including any noise present. Providing a measure of any noise present in the data set allows training to be terminated at a near optimal point. This is because an ANN will tend to fit useful data before any noise. Therefore, the Gamma test statistic  $\Gamma$  provides a MSE value at which training can be stopped (for example, an approximate target value of 13.45, shown in Figure 7.13, for the City Centre cluster).

### 7.7.4 Partitioning the vectors into training and test sets

Once the number of inputs required to model the output is known the data can be transformed to fit the optimal set-up. Using this set-up, an M-test is performed to establish whether the available number of vectors is sufficient to model any underlying function. An asymptotic level for the Gamma statistic (which approximates to the inherent noise of the output) indicates that there is sufficient data and provides a point where the data can be split into training and test vector sets. This is an important consideration as it allows the data set to be split into two, rather than the commonly practised three, parts (the training, validation and test set). Consequently, there is no need to set aside part of the data as a validation set, which is used to determine when continuing to train would result in over-fitting, thus allowing a higher proportion of data to be utilised during training (Wilson, *et al.* 2002). Thus, selection of the appropriate amount of data for modelling is confirmed at a point in advance of where the M-test reaches a stable level indicative of inherent noise (for example, an approximate volume of 300, shown in Figure 7.14, might be taken as the partitioning point).



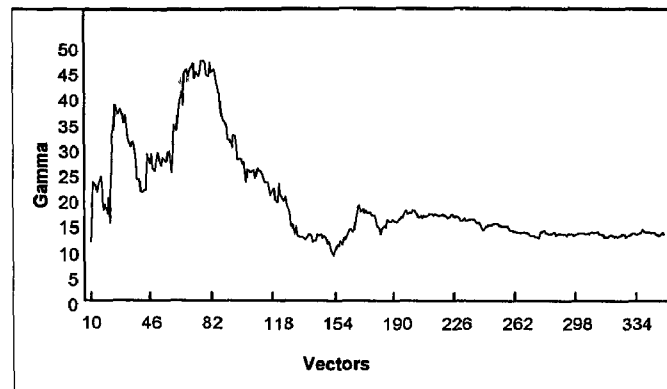


Figure 7.14 M-Test for the City Centre

Alternatively, a non-asymptotic level indicates that either there is insufficient data or a useful underlying function cannot be extrapolated from the data.

## 7.8 Experimental Work

ANN and comparative linear regression forecasting models were constructed using the GT, and compared to a 'random walk' (RW) model. The RW forecasts the change from  $t_i$  to  $t_{i+1}$  based upon the average change from one period to the next. For example, taking the known number of crimes for a Thursday the forecast for the following day is based upon the average observed change (over the entire time series) between Thursday and Friday.

### 7.8.1 Comparison between ANN, Linear Regression and Random Walk

As an example of the results obtained, two of the ANN models are discussed here, focusing upon two clusters analysed according to daily incident count.

#### Cluster Seven (residential)

The GT was used to determine window length (30), training (308) and test (28) vector partitioning, and noise estimate (8.01, or 32% of the range, a high value indicating a very chaotic data series). In addition, the gradient statistic estimate (0.0159) suggested that relatively few hidden nodes (10 in each of the two hidden layers) would be required to reach the estimated Gamma statistic. The resultant ANN, Regression and RW models resulted in error percentages of approximately 31.3%, 30.5% and 30.9% of the range, respectively.

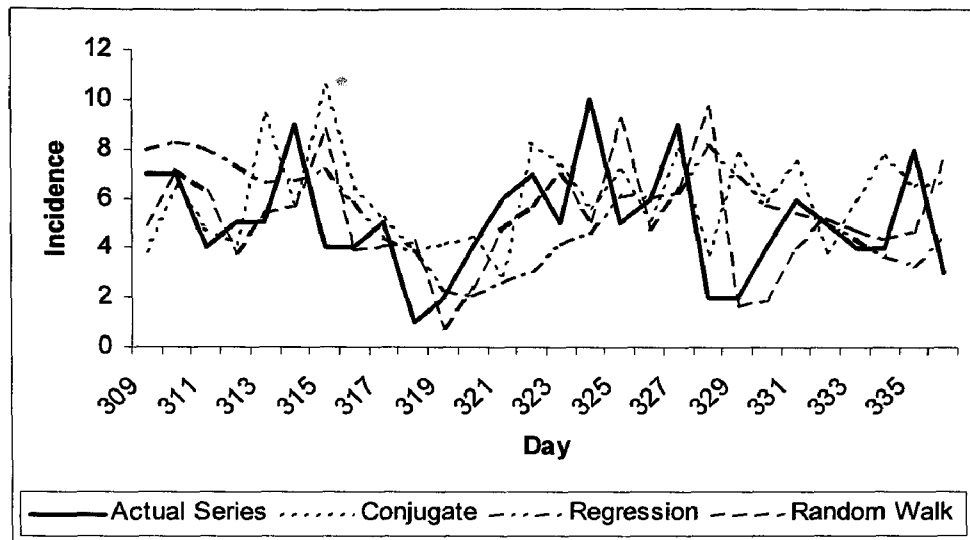


Figure 7.15 Incidence and forecast of violent crime (Cluster 7)

	Conjugate	Regression	Random Walk
Average	7.95	7.61	7.77
MSE			
Accuracy	31.3%	30.5%	30.9%
Min	1		
Max	10		
Range	9		

Table 7-2 Results for Cluster Seven

Cluster seven, which relates to a residential area with very few owner-occupiers, showed almost no general correlation between incident rate and weekday (illustrated in Figure 7.15). However, an increased tendency for crime towards weekends was noted, warranting a closer examination of other causal factors.

#### Cluster Four (City Centre)

The GT procedures were utilised to determine window length (13), training (330) and test (23) vector partitioning, and noise estimate (13.1365, or 27% of the range). In addition, the gradient statistic estimate (0.0159) suggested that relatively few hidden nodes (5 in each of the two hidden layers) would be required to reach the estimated Gamma statistic. The results generated by the resultant ANN, Regression and RW models (shown in Figure 7.16 and Table



7-3) produced error percentages of approximately 24.2%, 33.5% and 36.4% of the range, respectively.

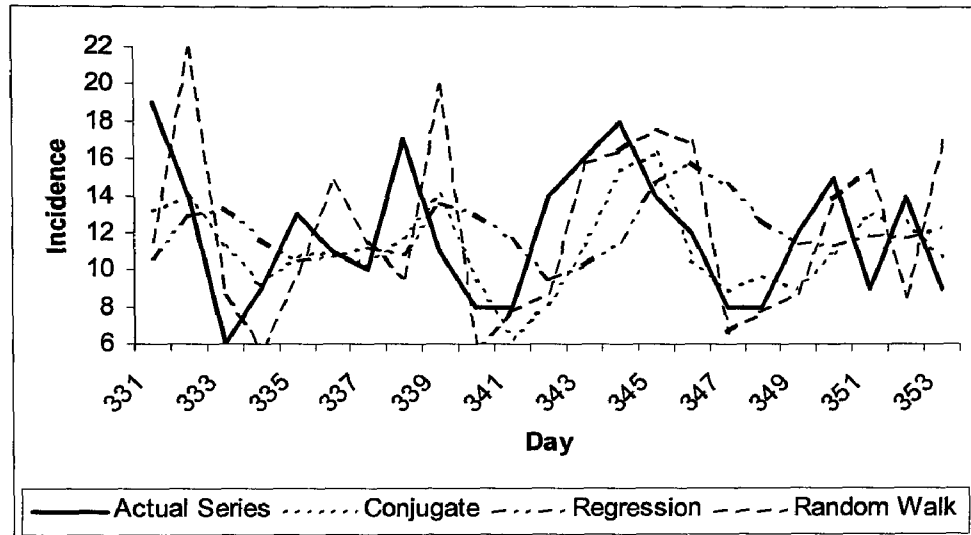


Figure 7.16 Incidence and forecast of violent crime (City Centre)

	Conjugate	Regression	Random Walk
Average	9.94	18.96	22.50
MSE			
Accuracy	24.2%	33.5%	36.4%
Min	6		
Max	19		
Range	3		

Table 7-3 Results for Cluster Four

Cluster four, which relates to the city centre, along with a concentrated collection of night clubs, public houses, public transport centre, and a sporting stadium, showed a higher incidence of crime at weekends, with peaks during times of known sporting events.

## 7.9 Discussion of results

The results concurred with expectations given the GT's output. Importantly, exceptional incidence levels that occur infrequently and appear as noise are excluded from the underlying model and account for a large portion of the error margin. Also, the utility of the GT was demonstrated as a pre-model evaluation technique. Moreover, the results show that the

methodology has the capacity to model the cause and effect relationship where one exists. The city centre (cluster four) offered the best predictive model using the ANN, cluster seven (residential area) generating relatively poor models for the ANN, Regression and RW.

### 7.10 Summary and Conclusions

The work presented here introduces the first stage (summarised in Figure 7.2) in the development of a computerised system designed to facilitate accurate crime incident forecasting by focusing upon geographical areas of concern that may well transcend traditional policing boundaries. The research focuses upon the development of a practical solution, for use in an operational policing environment, which ameliorates the deficiencies of these rigid boundaries and moves towards a more dynamic methodology. The computerised procedure utilises a geographical crime incidence scanning algorithm to identify clusters with relatively high levels of crime (hot-spots). These clusters provide sufficient data that can be analysed using the GT procedures, assessing fitness and required configuration (for example, quantity of data required and number of inputs) for predictive modelling. Using the outputs from the GT, two techniques were implemented (ANN and Regression). The ANN generally exhibited a superior capacity to model the trends within each cluster. RW was used as a naïve forecasting method, the results demonstrating a comparable forecasting accuracy to the other techniques for cluster seven (residential area) where the GT indicated a chaotic data series.

Future developments should include the modelling of more detailed scenarios to facilitate prediction based upon selected input criteria. Thus, for example, the impact upon the region of, say, a forthcoming public holiday, where the weather is predicted to be warm, could be evaluated. The objective here was to extract an underlying, generalised, model of crime incidence. However, specific localities might best be modelled independently of the other data at specific times of the year (for example, sporting events that generate exceptionally high crime *spikes* that fall out of a generalised model). These spikes could be extracted and treated as a separate modelling exercise, given that there is both sufficient data of high quality. Alternatively, a statistical analysis of exceptional events could provide an estimate of the change against normal levels. This could then be used to modify the incidence count accordingly. These two differing approaches should form the basis for future experimentation.

## 8 CONCLUSIONS AND FUTURE WORK

*This Chapter provides a synopsis of the research and discusses the key contributions to knowledge. The Chapter concludes with a discussion on how the research might be progressed and provides some suggestions for future work.*

---

### 8.1 Introduction

The project research objectives were finalised following a rigorous ethical appeasement process taking account of both available data and imposed restrictions on their use. The research endeavoured to:

- (1) Evaluate the current use and uptake of computerised mapping technologies for operational and strategic incident mapping;
- (2) Develop a GIS that is descriptive of crime and disorder issues integrating all relevant data sets;
- (3) Implement and evaluate exploratory and predictive techniques for use on crime and incident databases;
- (4) Develop a GIS for visualisation of geospatial information retrieved from an analytical engine comprising Artificial Intelligence paradigms.

Preceding Chapters have presented research that collectively has addressed each of these objectives. The following section discusses each objective in turn and provides evidence of fulfilment by drawing on the outcomes from the research programme.

(1) Evaluate the current use and uptake of computerised mapping technologies for operational and strategic incident mapping.

In order to evaluate the uptake of computerised mapping technologies, three national surveys were compiled and sent to each of the three emergency services (police, fire and ambulance). Posing regional-wide questions, current and perceived future engagement with computerised mapping was evaluated.

The survey generated a 52% (police), 54% (ambulance) and 54% (fire) response rate. Total UK coverage was not achieved reasons for which, as indicated by some non-respondents, included either lack of time or a reluctance to disclose certain intelligence information to questions posed by the survey tool.

The results indicated both a growing influence and positive sentiment toward the use and support of computerised mapping technologies. However, for several emergency services, numerous technical and political barriers currently impede its full exploitation. One finding was the degree to which each emergency service has procured and developed their own mapping programmes. This was highlighted by the numerous software combinations that are currently employed. This finding was of importance when considering the incorporation of additional techniques such as those presented in this thesis. In such cases issues of flexibility are paramount.

Working with two local Crime and Disorder Reduction Partnerships (Cardiff and Barry) further reinforced the issues highlighted by the three national surveys. This provided an insight into other agencies' engagement with spatial information beyond those of the emergency services. The most prominent of these was the central drive to utilise spatial information to inform strategic activities. In many cases this requirement placed new demands on partners, their IT infrastructures and their data collection practices.

(2) Develop a GIS that is descriptive of crime and disorder issues integrating all available and relevant data sets.

To collate the necessary spatial and non-spatial data sets required a rigorous and lengthy data procurement and ethical appeasement process. This resulted in the acquisition of two key data sets describing crime and injury as recorded by the police and A&E services respectively. All data was subsequently linked and visualised in conjunction with one

another using both cartographic and graphic representations to incorporate socio-economic, calendar events and meteorological information that were descriptive of the Cardiff region.

In addition to the creation of the GIS described above, work that involved the assembly of a similar series of data for completion of a local crime audit was undertaken. Working under the caveat of the Crime and Disorder Act (1998), in conjunction with the local Crime and Disorder Reduction Partnership, a broader series of crime and crime-related data was assembled. This was facilitated through the production of a data provision proforma stipulating minimum data requirements to permit a spatial analysis of partner's data.

(3) Implement and evaluate exploratory and predictive techniques for use on crime and incident databases.

Combinations of exploratory and predictive techniques were applied to police (Command and Control) and A&E (assault) data that were coupled with additional contextual information (for example, calendar events and holiday periods).

Exploratory

Common statistical and temporal geographical methods were applied to explore patterning, relationships and trends within crime, injury and contextual data sources. The selection of appropriate techniques was largely controlled by the content of the data. The A&E data contained temporal descriptors (i.e. date) coupled with contextual parameters (i.e. age and gender) to which a range of count, correlation and cross tabulation techniques were used on the entire and subsets of the time series. Police data, on the other hand, included locational information but no information on age or gender of the individual(s) involved. Exploratory analysis was therefore designed to target geo-temporal parameters. This involved the application of a range of cartographic and geo-statistical techniques upon a selected subset of incident types.

The comparison of A&E against police data (specifically violence against the person) offered an insight into the approximate level of underreporting within police records. Using simple counts and percentages over various granularities of time (i.e. daily and monthly). The A&E data consistently reported higher volumes of violence (40.9% difference) throughout the coverage period. The difference became more pronounced during weekend periods (59.2% higher than police recorded violence).

The animation tool was developed to offer an automated method by which data snapshots could be aggregated to create a single movie sequence. A viewer was developed to offer both a single and dual view port capability. This enabled either a single or a pair of animated sequences to be played. Evidence from the animation of density mapping indicated that hotspots were neither spatially nor temporally static. Local flow analysis was developed to complement the animation tool and identify these hotspot movements. By locating the centres of hotspots and tracking them over a time sequence, the flow analysis graphically represented the spatio-temporal lifecycles of these hotspots.

#### Predictive

A novel predictive technique was applied to the geo-temporal forecasting of crime. The technique used a geographical scanning technique to stratify the data into a series of geographical sub units (clusters). The concept of this technique made the assumption that crimes of the same type in the same area held similar characteristics and therefore would form the basis of reliable predictive models. Each cluster then formed the basis for a separate predictive model. The Gamma Test was applied to the data in each cluster to assess their applicability to be modelled. The Gamma Test assessed the amount of data that could not be explained by a smooth transformation (intercept), and an indication of complexity (gradient). Using both indicators decisions could be made, either to progress and train an ANN, or to geographically redefine the input data to improve its Gamma statistic.

- (4) Develop a GIS for visualisation of geospatial information retrieved from an analytical engine comprising Artificial Intelligence paradigms.

The CLAP interface was used to create the first tangible link between a geographical interface to display incident data and ANNs trained from the results of a cluster analysis. The CLAP was the first stage in the development of a flexible tool that could handle the diversity of operational configurations (Chapter Three) and offered a tool to improve the efficiency of current resource allocation practices. CLAP was developed as a self-contained piece of software, therefore placing limited requirements upon existing databases and mapping packages, and increasing applicability for a widespread implementation. A central element to CLAP is independence from traditional policing boundaries, the clustering and predictive outputs computed for areas that typically transcend these. The technique permitted a more accurate depiction of crime volumes

and, coupled with a prediction for these areas, offers the potential for an improved deployment of resources.

## **8.2 Concluding remarks**

Aggregating the various facets of the research presented, a series of key conclusions can be drawn:

- Analysing the geography of crime has become an established approach for developing an understanding of criminal dynamics. GIS forms an essential feature of this approach;
- A truer depiction of crime in an area is achieved where police data is supplemented with additional crime and contextual data. However, this is subject to both ethical and technical constraints;
- Crime is neither static in time nor space. Animation tools provide a useful insight into their dynamics;
- The temporal geography of crime can be predicted with a level of success where training data is sufficient. A series of models, each dedicated to a specific geographical subset of data, provides the greatest level of accuracy.

## **8.3 Contribution to knowledge**

The research presented in this thesis broadly encompasses the disciplines of crime auditing, geographical information systems and crime prediction, and has made the following contributions to each of these fields:

### Crime Auditing

HASCADE was successfully implemented as a geographical method to holistically analyse multi Agency crime and disorder data. The advantage of HASCADE over other methods is its ability to derive a concise picture (a single cartographic output) depicting crime and vulnerability across a region. The output can then be used as a basis from which strategy can be developed. By differentiating between crimes and vulnerabilities, different interventions can be considered.

### Geographical Information Systems (GIS)

Within the field of GIS, the research has resulted in the development of an animation tool that automates the creation of an animated sequence from a series of input maps. This tool provides a valuable insight into the dynamics of various data sets over space and time. This technique is supplemented with local flow analysis where movements of crime were analysed.

Examining the life cycles of hotspots (area and centres), local flow analysis identified that certain crime types exhibited well-defined space-time movements.

#### Crime Prediction

Combining the fields of both GIS and Artificial Intelligence (specifically neural networks) a methodology was implemented using the CLAP interface to facilitate the geo-temporal prediction of crime. Using a geographical scanning algorithm, clusters of crime were identified. The methodology employs the Gamma Test to evaluate the quality of each cluster's data for predictive modelling. In addition, the Gamma Test was used to determine appropriate neural network architectures.

#### **8.4 Potential avenues for future research**

The research presented in this thesis encompasses a variety of agencies and techniques. However, the main focus has been with a geographically orientated exploration and analysis of crime and related information. The following section discusses a series of potential future research avenues that have arisen from the research.

- Data collection

The quality of results from any analytical technique applied to any data is subject to the accuracy of that data. This coupled with the motivation to place a greater emphasis on a geographically orientated approach based upon individually referenced records derived from the police systems (in addition to the other ES databases), places reliance upon robust and reliable spatial and temporal identifiers. Data collection methodologies are therefore crucial in controlling data reliability.

Existing methods of collecting incident data involves the reporting officer making a scene of crime report consisting of paper details on return to the station. A separate operator enters this sheet of information onto a database. Due to the number of stages, this process has the potential to create errors in the data. Subsequent analysis of the data may produce misleading results. Errors in the reporting process and their impact on spatial and temporal analysis have been noted by previous studies (Ekblom 1988; Ratcliffe and McCullagh 1998; Bowers and Hirschfield 1999). Table 8-1 details some typical errors:

The advent of PalmTop technology interfaced with a Global Positioning System (GPS) offers potential to avoid certain errors (for example, locational and temporal parameters) that currently affect the integrity of Police databases. The ability to standardise the data



collection process will in turn improve the potential for a superior analysis. GPSs have been implemented in a range of policing applications in both the US and UK (Chapter Three). As wireless PalmTop technology become more widely established the addition of a GPS offers a powerful data capture tool to address some of the issues affecting the integrity of police databases.

ERROR TYPE	EXAMPLE
Spelling omissions	Misspelt names and addresses
Notation errors	Incorrect use of abbreviations such as male, female, 1,3
Inconsistent notation	Mixture of notation styles to denote the same attribute (for example, Ave, Avenue, AVE).
Transfer / Interpretation errors	Incorrect interpretation of reporting officers notes transcribed onto the database
Inaccurate description of incident location / date / time	Imprecise temporal and spatial representation of an incident
Duplication	J Higgins, JHIGGINS

Table 8-1 Typical errors found in Police database records

- Acquisition of additional data sources

Crime is a complex phenomena and, although a level of success was achieved through time series modelling (Chapter Seven), greater potential for explanation lies in the use of multiple explanatory variables (Chapters Six).

Explanatory variables used in this research included a variety of calendar events (for example, sports and holidays periods), and meteorological and socio economic indicators. Calendar and meteorological parameters have an impact upon the movement of individuals, for example, the day of an international sporting event where the city centre can witness the influx of 80,000 individuals (seating capacity of the stadium). This movement of individuals was identified using animation and quantitative techniques tracking flows of criminal activity (Chapter Six). Thus far the analysis has been founded upon locales of crime, however, a supportive piece of research would be an investigation

of use of these areas by individuals (i.e. pedestrians) not resulting in a crime being committed.

The morphology of Cardiff city centre with the river Taff to the west and large parkland and main road to the north-west are such as to constrain the number of input and output opportunities (Figure 8-1). Two main railway stations and the central bus station provide access at the southern and south-eastern fringes. Internal to the centre are a series of bus stops located along St. May's street and several taxi ranks.

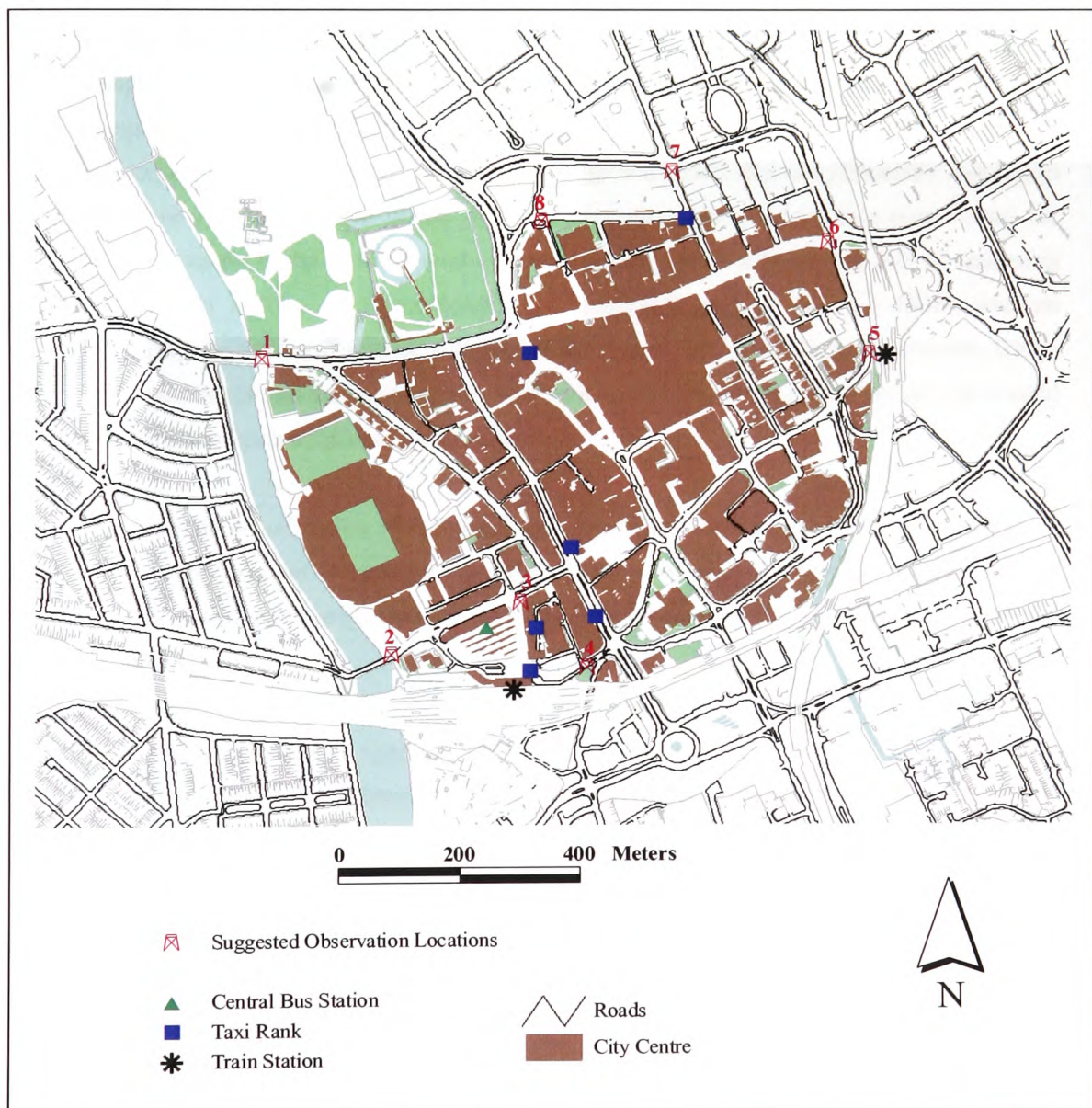


Figure 8-1 Cardiff city centre

Figure 8-1 illustrates that a series of eight observation points would provide sufficient information on pedestrian movement to and from the city centre. Using automated

counting of CCTV coverage represents the least resource intensive option to derive the necessary data. Counts of entrants and leavers from the city centre could then be made at intervals over 24-hour cycles and a seven-day period. Based upon evidence presented in Chapter Six, the weekly cycles were the most dominant, thus a seven day period should provide sufficient information. This could then be evaluated against calendar events and prevailing weather conditions coupled with their relationship to locales, timings, volume and type of criminal activity.

- Transparent analysis tools

Chapter Three highlighted that, although there are some common approaches in the use of computerised mapping, numerous disparities exist. This is coupled with the general limitations of standard mapping environments to offer tools that can efficiently respond to the gamete of operational questions. Building upon the previous two suggestions for future research is the development of supplementary tools that are capable of analysing the data, to address such questions. The CrimeStat package (Levine 2002), with its capability to read and output a range of file types, offers a blueprint towards which additional tools should be founded. This would account for the diversity of operational software configurations noted previously.

Focusing upon animation (developing that introduced in Chapter Five) as one area of analysis that has yet to be exploited fully, future work should target the development of additional software tool(s) that are capable of exploring patterns depicted in the animated outputs.

- Integrated predictive solutions

A common thread throughout each of the suggestions for future work is the endeavour to achieve a better understanding of criminal activity. At this point interventions can then be targeted to suit the identified issues. The ability to computerise the data collection process (using handheld terminal interfaced with GPS) presents an opportunity to conduct the entire process from collection, analysis and presentation of the results seamlessly, without the requirement to translate paper documents into electronic records. This would have the effect of both reducing the potential for error and increasing the information processing efficiency. Using handheld terminals connected to a wireless network, pertinent information could be transmitted to patrolling officers. Information could be represented either as a text based message or cartographic illustrations within which their and nearby officers current localities are plotted.

Coupled with the data collection element is the analysis component through which patterning and predictive estimates can be made. Chapter Seven presented an approach using the CLAP interface to predictively model crime data. The natural progression of these techniques is to develop a scenario based prediction founded upon the proven ANN methodology. Using a range of data inputs (for example, pedestrian flow, metrological and sporting event information) a computation of predictive estimates in the form of a risk map could be produced. The cartographic output would be representative of the input scenario and based upon trained models common to the selected scenario. The results of this scenario could then be disseminated to police officers through handheld terminals.

- Clustering using ANN techniques

Simple statistical methods reveal that levels of crime vary considerably depending on the time of day, the day of the week (Van Koppen and Jansen 1999), prevailing weather conditions (Harries, *et al.* 1984; Anderson 1987) and a legion of other influencing factors, such as unemployment fluctuations (Falk 1952). However, what is less clear is the degree to which each contributing factor affects crime levels. ANNs offer one possible solution for determining the cause and effect relationship.

Chapter Seven presented a methodology founded upon the geographical stratification of crime incidence into a series of geographical sub-units for which individual predictive models were implemented. Using a moving window approach this methodology removed any additional contextual variables from the predictive model and is founded solely upon a continuous time series. ANN, and specifically Kohonen Self Organising Maps (Kohonen 1984) offer a technique to automatically stratify the heterogeneous input data into a series of homogeneous sub-units.

In the case of crime forecasting, the input vectors could be numerically coded representations of the suspected casual factors and the incident count respectively. For example, the input vectors are of the form (Location, Day, Time, Weather, Sports Event, School Holiday). A typical vector could be (Cluster 1, Saturday, 15:00, wet, rugby, no) which needs to be coded to its numeric equivalent, for example, (1, 0.7, 0.8, 1, 0, 0). In order to “train” correctly there must be a relationship – albeit unknown - between those deemed to be causal factors and crime levels. Moreover, an extensive training set covering the whole spectrum of possible input and output vectors would be required.

The objective would be to identify clusters within the input vectors that are typically difficult to achieve using simple sort procedures. These clusters can then be extracted and inspected to form the basis of a rule abduction exercise to discern common characteristics describing a cluster (Figure 8-2).

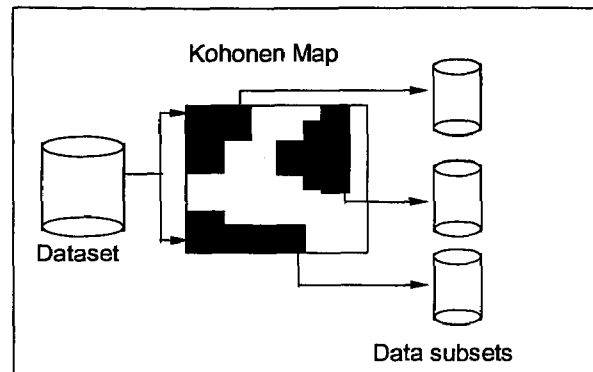


Figure 8-2 Using a KSOM to partition a dataset

- Implementation and usability testing

The series of techniques and methodologies presented throughout this thesis have in the main been external to any operational environments. The next step in their development should be to compile a series of tools for deployment to collaborating emergency services for usability evaluation.

## 8.5 Final Remarks

Computerised crime mapping and analysis using Geographical Information Systems will increase in importance. As these systems become more established they will be supplemented by additional analytical techniques seeking to intelligently derive greater explanation from the recorded data.

The development of a predictive geographical environment will form one additional tool. When the techniques presented in this thesis are suitably developed into deployable solutions, their added value to operational decision-making can be evaluated. At that point it will be possible to respond to the “what”, “where” and “when” questions facilitating the deployment of suitable preventative measures.

## REFERENCES

- Alber (1987). "The national science foundation national centre for geographic information and analysis." International Journal of Geographical Information Systems 3(4): 303-326.
- Anderson, C. A. (1987). "Temperature and aggression: Effect on quarterly, yearly and city rates of violent and non-violent crime." Journal of Personality and Social Psychology 52: 1161-1173.
- Anderson, C. A. (1989). "Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence." Psychological Bulletin 106: 74-96.
- Anselin, L. (1995). "Local indicators of spatial association - LISA." Geographical Analysis 27(2): 93-115.
- Anselin, L., J. Cohen, D. Cook, W. Gorr and G. Tita (2000). "Spatial analyses of crime." Criminal Justice - Measurement and Analysis of Crime and Justice 4.
- Badiru, A., J. M. Karasz and R. T. Holloway (1988). "AREST: Armed robbery eidetic suspect typing expert system." Journal of Police Science and Administration 16: 210-216.
- Bailey, T. C. (1994). A review of statistical spatial analysis in geographical information systems. Spatial Analysis and GIS. A. S. Fotheringham and P. Rogerson. London, Taylor and Francis: 13-44.
- Baldwin, J. (1975). "British areal studies of crime: An assessment." British Journal of Criminology 15(3): 211-227.
- Barr, R. and K. Pease, Eds. (1990). Crime placement, displacement, and deflection. Crime and justice: A review of research. Chicago, University of Chicago Press.
- Barry, D. (1991). Fleet management makes advances with digital mapping technology. GIS World. 4: 74-77.
- Batty, M. (1995). Animating urban dynamics: Using GIS to visualise space-time histories and predictions of cities from the industrial and post-industrial eras. GIS Research UK, Newcastle.44-45.
- Bottomley, A. K. and C. A. Coleman (1976). "Criminal statistics: the police role in the discovery and detection of crime." International Journal of Criminology and Penology 4: 33-58.
- Bowen Thomson, B. and J. Corcoran (2002). Cardiff crime and disorder audit 2001 and strategy 2002-2005. Cardiff, Cardiff Community Safety Partnership.
- Bowers, K. and A. Hirschfield (1999). "Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems." International Journal of Geographical Information Science 13(2): 159-184.

- Brahan, J. W., K. P. Lam, H. Chan and W. Leung (1998). "AICAMS: Artificial intelligence crime analysis and management system." Knowledge Based Systems 11: 355-361.
- Brantingham, P. J. and P. L. Brantingham Eds. (1981). Environmental criminology. Beverly Hills, Sage.
- Brantingham, P. L. and P. J. Brantingham (1993). Environment, routine and situation: Toward a pattern theory of crime. Routine activity and rational choice. R. V. Clarke and M. Felson. London, Transaction Publishers. 5: 259-294.
- Brown, D.E. (1999). Predictive models for law enforcement. MAPS Grant Portfolio. <http://nij.ncjrs.org/portfolio/> Last accessed 04/04/2003
- Brown, D. E. (1998). The regional crime analysis program (RECAP): A framework for mining data to catch criminals. IEEE International Conference: Systems Man and Cybernetics Society.2848-2853
- Brown, P. J. B. and P. W. J. Batey (1994). Characteristics of super profiles lifestyles and target markets: Index tables, pen pictures and geographical distribution, Liverpool UPERRL, Department of Civic Design, Liverpool University.
- Brunsdon, C., A. S. Fotheringham and M. Charlton (2002). "Geographically weighted summary statistics - a framework for localised exploratory data analysis." Computers, Environments and Urban Systems 26: 501-524.
- Burt, C. (1925). The young delinquent. London, *London University Press*.
- Carcach, C. (2000). The spatial analysis of ambulance calls for drug overdose in Adelaide. Searching for a link between drug use and property crime. Crime Mapping: Adding Value to Crime Prevention and Control, Adelaide
- Castle, I. M. and E. Gittus (1957). "The distribution of the social defects in Liverpool." Sociological Review 5: 43-64.
- Cattini, M. (1997). "London calling: ambulance services get better with GIS." Mapping Awareness 11(3): 35-37.
- Chainey, S. and S. Reid, Eds. (2002). When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime. Socio-economic applications in geographical information science, Taylor & Francis.
- Cheatwood, D. (1995). "The effects of weather on homicide." Journal of Quantitative Criminology 11: 51-70.
- Clark, P. J. and F. C. Evans (1954). "Distance to the nearest neighbour as a measure of spatial relationships in populations." Ecology 35: 445-453.
- Clarke, R. V. and M. Felson, Eds. (1993). Routine activity and rational choice. Advances in Criminological Theory. New Jersey, Transaction.
- Coady, W. F. (1987). "Investigating with APES." Security Management 31: 67-70.



- Cohen, L. E. and M. Felson (1979). "Social change and crime rate change: A routine activities approach." *American Sociological Review* 44: 588-608.
- Cornish, D. B. and R. V. Clarke, Eds. (1986). The reasoning criminal: Rational choice perspectives on offending. New York, Springer-Verlag.
- Curtin, L., N. Tilley and K. Pease (2001). Developing crime reduction plans: Some examples from the Reducing Burglary Initiative. London, Home Office.
- Derekenaris, G., J. Garofalakis, C. Makris, J. Prentzas, S. Sioutas and A. Tsakalidis (2000). "Information system for the effective management of ambulances." Proceedings of the IEEE Symposium on Computer-Based Medical Systems: 269-273.
- Dorling, D. and S. Openshaw (1991). Experiments using computer animation to visualise space-time patterns. Proceedings of Second International Conference on Computers in Urban Planning and Management, Oxford.391-406
- Durrant, P. J. (2001). winGamma™: A non-linear data analysis and modelling tool with applications to flood prediction. Department of Computer Science. Cardiff, Cardiff University.
- ECRI-(Environmental-Criminology-Research-Inc) (1999). The RIGEL system. <http://www.ecricanada.com/Rigel.htm> Last accessed 04/04/2003
- Ekblom, P. (1988). Getting the best out of crime analysis, Police Research Group.
- ERA.Technology.Ltd (1999). Case studies at ERA: Out-smarting the criminals. <http://www.era.co.uk/case/cs9907.htm> Last accessed 05/06/2001
- ESRI (1999). Crime analysis application extension for ArcView GIS, Environmental Research Institute Inc. (ESRI).
- Fairwater, B. (2000). Personal data: Issues of ethics and regulation, ESRC. <http://www.ccsr.cse.dmu.ac.uk/conferences/ESRC/PersonalData.html> Last accessed 04/04/2003
- Falk, J. J. (1952). "The influence of the seasons on the crime rate." *Journal of Crime and Law Criminology* 43: 199-213.
- Farrington, D. P. (2002). Developmental criminology and risk-focussed prevention. The Oxford handbook of criminology. M. Maguire, R. Morgan and R. Reiner. Oxford, Clarendon Press.
- Felson, M. (2002). Situational crime prevention workshop. Personnel communication with J. Corcoran. London.
- Felson, M. and R. V. Clarke (1998). Opportunity makes the thief: Practical theory for crime prevention. London, Home Office: Police Research Series Paper 98.
- Fletcher, J. (1849). "Moral Statistics of England and Wales." *Journal of the Royal Statistics Society of London* 12: 151-181 : 189-335.



- Fotheringham, S. A., C. Brunsdon and A. Charlton (2002). Geographically weighted regression - the analysis of spatially varying relationships. Chichester, John Wiley & Sons Ltd.
- Fotheringham, S. A. and P. Rogerson (1993). "GIS and spatial analytical problems." International Journal of Geographical Information Systems 7(1): 3-19.
- Francis, S. H. (1971). Prediction methods in criminology, *Home Office Research Studies*.
- Gabor, T. (1990). "Crime displacement and situational prevention: Towards the development of some principles." Canadian Journal of Criminology 32: 41-74.
- Geary, R. (1954). "The contiguity ratio and statistical mapping." The Incorporated Statistician 5: 115-145.
- Georges, D. E. (1978). The geography of crime and violence: a spatial and ecological perspective.
- Getis, A. and J. K. Ord (1992). "The analysis of spatial association by use of distance statistics." Geographical Analysis 24: 189-206.
- GIS-Newslink (1993). "Emergency mapping response implemented." GIS World 6(7): 11.
- Glueck, S. E. (1960). Predicting delinquency and crime. Cambridge, Massachusetts, Harvard University Press.
- Goodchild, M. F. (1987). "A spatial analytical perspective on geographical information systems." International Journal of Geographical Information Systems 1(4): 327-334.
- Goodwin, V. and J. P. Shepherd (2000). "The development of an assault questionnaire to allow A&E departments to contribute to crime and disorder act local crime audits." Journal of Accident Emergency Medicine 17: 1-2.
- Gorr, W., A. Olligschlaeger and Y. Thompson (2000). Assessment of crime forecasting accuracy for deployment of police. Pittsburgh, Carnegie Mellon University.
- Gorr, W. and A. Olligschlaeger (1998). Crime hot spot forecasting: Modelling and comparative evaluation. MAPS Grant Portfolio. <http://nij.ncjrs.org/portfolio/> Last accessed 04/04/2003
- Graham, J., P. Ekblom and K. Pease (1998). Reducing offending: An assessment of research evidence on the ways of dealing with offending behaviour. London, Home Office: Research study 187.
- Guerry, A. M. (1833). "Essai sur la statistique morale de la France." Westminster Review 18(357).
- Harries, K. D. (1974). The geography of crime and justice. New York, McGraw-Hill.
- Harries, K. D. (1975). "Rejoinder to R. Peet: "The geography of crime: A political critique." The Professional Geographer 27(3): 280-282.

- Harries, K. D., S. J. Stadler and R. T. Zdorkowski (1984). "Seasonality and assault: explorations in inter-neighbourhood variation, Dallas 1980." Annals of the Association of American Geographers 74(4): 590-604.
- Herbert, D. T. (1976). "The study of delinquency areas: a social geographical approach." Institute of British Geographers Transactions 1(4): 472-492.
- Herbert, D. T. (1982). The geography of crime, Longman.
- Home Office (2002). Crime reduction toolkits. <http://www.crimereduction.gov.uk/toolkits/> Last accessed 04/04/2003
- Home Office (2000a). Review of crime statistics: A discussion document. <http://www.homeoffice.gov.uk/crimprev/crimstco.htm> Last accessed 07/06/2001
- Home Office (2000b). Criminal statistics - England and Wales. London, Home-Office.
- Hornik, K. (1991). "Approximation capabilities of multi-layer feed-forward networks." Neural Networks 4: 251-257.
- Hough, M. and N. Tilley (1998). Auditing crime and disorder - Guidance for local partnerships. London, Home Office.
- Hylleberg, S., Ed. (1995). Modelling seasonality. Oxford, Oxford University Press.
- Icove, D. J. (1986). "Automated crime profiling." Law enforcement Bulletin 55: 27-30.
- Illinois-Criminal-Justice-Information-Authority (1988). Spatial and Temporal Analysis of Crime (STAC). Chicago.
- ISL (1999). Data mining case studies: West Midlands Police - crime detection. <http://www.isl.co.uk/clementine/wmpolice.html> Last accessed 16/04/2000
- Jones, C. D. (1934). Social survey of Merseyside, Liverpool University Press.
- Jones, J. A., D. Evans, S. Margetts and P. J. Durrant (2002). The gamma test. Heuristic and optimisation for knowledge discovery. R. Sarker, H. Abbass and C. Newton. Hershey, Idea Group Publishing.
- Kershaw, C., T. Budd, G. Kinshott, J. Mattinson, P. Mayhew and A. Myhill (2000). The 2000 British crime survey. London, Home Office.
- Kohonen, T. (1984). Self-Organisation and associative memory. New York, Springer-Verlag.
- Lee, Y. (1975). "A rejoinder to "The geography of crime: A political critique"." The Professional Geographer 27(3): 284-285.
- Levine, N. (1999). CrimeStat. Annandale, National Institute of Justice.
- Levine, N. (2002). CrimeStat. Washington, DC, National Institute of Justice.

- Maguire, M. and H. Nettleton (2002). Reducing alcohol-related violence and disorder: An evaluation of the 'TASC' project. London, Home Office.
- Maguire, T. J. (1996). Crime and community safety in the Vale. Barry, Vale County Borough Council.
- Mamalian, C. A. and N. G. LaVigne (1999). The use of computerised crime mapping by law enforcement: Survey results, *National Institute of Justice Research Preview*. <http://www.ncjrs.org/pdffiles1/fs000237.pdf> Last accessed 04/04/2003
- Mannheim, H. and L. T. Wilkins (1955). Prediction methods in relation to Borstal training. London, *H.M.S.O.*
- MAPS (2003). CRIMEMAP Listserv, National Institute of Justice. <http://www.ojp.usdoj.gov/nij/maps/> Last accessed 04/04/2003
- Mayhew, H. (1862). *London Labour and the London Poor*. London, Griffin-Bohn.
- McCabe, S. and F. Sutcliffe (1978). *Defining a crime: A study of police decisions*. Oxford, *Oxford:Blackwell*.
- Moran, P. A. P. (1950). "Notes on continuous stochastic phenomena." *Biometrika* 37: 17-23.
- Ohlin, L. E. and O. D. Duncan (1949). "The efficiency of prediction in criminology." *American Journal of Sociology*: LIV 5.
- Olligschlaeger, A. and W. Gorr (1997). Artificial neural networks and crime mapping. *Crime mapping and crime prevention*. D. Weisburn and J. T. McEwen. New York, Criminal Justice Press. 8: 313-347.
- Openshaw, S., M. Charlton and A. Craft (1987). "A Mark I geographical analysis machine for the automated analysis of point data series." *International Journal of Geographical Information Systems* 1(4): 335-358.
- Openshaw, S., A. Cross, A. Charlton and C. Brunsdon (1990). *Lessons learnt from a post mortem of a failed GIS. 2nd National Conference and Exhibition of the AGI, Brighton*
- Openshaw, S., D. Waugh and C. A. (1994). Some ideas about the use of animation as a spatial analysis tool. *Visualisation in GIS*. H. W. U. Hearnshaw, D.J., John Wiley & Sons Ltd.
- Ordnance-Survey (2001). Code-Point® Data. Southampton, UK.
- Pauly, G. A., J. T. McEwen and S. Finch (1967). Computer mapping - A new technique in crime analysis. *Law Enforcement Science and Technology*. S. A. Yefsky. New York, Thompson Book Co. 1.
- Peet, R. (1975). "The geography of crime: A political critique." *The Professional Geographer* 27(3): 277-280.
- Peters, J. and G. B. Hall (1999). "Assessment of ambulance response performance using a geographic information system." *Social Science and Medicine* 49(11): 1551-1566.

- Phillips, C., Considine, M., Lewis, R. (2000). A review of audits and strategies produced by crime and disorder partnerships in 1999. London, Home Office.
- Phillips, P. D. (1972). "A prologue to the geography of crime." Proceedings, Association of American Geographers 4: 86-91.
- Phillips, P. D. (1975). "Radical theory, relevance and the geography of crime." The Professional Geographer 27(3): 283-284.
- Police-Foundation (2000). Users' guide to mapping software for police agencies. Washington, D.C., Crime Mapping Laboratory.
- Police-Foundation (2002). Crime Mapping News. Washington, D.C., Police Foundation.
- Povey, D., J. Prime and P. Taylor (1997). Notifiable offences: England and Wales. London, HMSO.
- Pyle, G. F. (1974). The spatial dynamics of crime. Chicago, Department of Geography, University of Chicago: *Research Paper 159*.
- Quêtelet, A. (1842). A treatise on man. Edinburgh, Chambers.
- Ratcliffe, J. H. (2000a). "Implementing and integrating crime mapping into a police intelligence environment." International Journal of Police Science and Management 2(4): 313-323.
- Ratcliffe, J. H. (2000b). Visualising crime hotspots and making sense of high volume crime. AIC conference: Crime Mapping: Adding value to crime, Adelaide, Australia.
- Ratcliffe, J. H. (2001). "Chasing ghosts? Police perception of high crime areas." British Journal of Criminology 41: 330-341.
- Ratcliffe, J. H. (2002). "Aoristic signatures and the spatio-temporal analysis of high volume crime patterns." Journal of Quantitative Criminology 18(1): 23-43.
- Ratcliffe, J. H. and M. J. McCullagh (1998). "Identifying repeat victimisation with GIS." British Journal of Criminology 38(4): 651-662.
- Ratcliffe, J. H. and M. J. McCullagh (1999). "Hotbeds of crime and the search for spatial accuracy." Journal of Geographical Systems 1: 385-398.
- Ratledge, E. C. and J. E. Jacoby (1989). Handbook on artificial intelligence and expert systems in Law Enforcement. New York, Greenwood Press.
- Reynolds, C and Pedroza, J. (1998). Fire cover modelling for brigades. London, Home Office.
- Robinson, W. S. (1950). "Ecological correlations and the behaviour of individuals." American Sociological Review 15: 351-357.
- Rogerson, P. (1998). Detection and prediction of geographical changes in crime rates. MAPS Grant Portfolio. <http://nij.ncjrs.org/portfolio/> Last accessed 04/04/2003.

- Rossmo, D. K. (2000). Geographic profiling. Boca Raton, CRC Press.
- Scott, P. (1972). "The spatial analysis of crime and delinquency." Australian Geographical Studies 10: 1-18.
- Shaw, C. R. and H. D. McKay (1969). Juvenile delinquency and urban areas. Chicago, Chicago University Press.
- Shaw, C. R., H. D. McKay, F. Zorbaugh and L. S. Cottrell (1929). Delinquency areas. Chicago, *University of Chicago Press*.
- Shepherd, J., M. Shapland and C. Scully (1989). "Recording by the police of violent offences; An accident and emergency department perspective." Medicine, Science and the Law 29: 251-257.
- Shepherd, J. P. (1990). "Violent crime in Bristol." The British Journal of Criminology 30(3): 289-305.
- Shepherd, J. P., M. A. Ali and A. O. Hughes (1993). "Trends in urban violence: a comparison of accident department and police records." Journal of the Royal Society of Medicine 86: 289-304.
- Simmons, J. (2002). Crime in England and Wales 2001/2002. London, Home Office.
- Sivarajasingham, V. and J. P. Shepherd (2001). "Trends in community violence in England and Wales 1995-1998: an accident and emergency department perspective." Journal of Emergency Medicine 18: 105-109.
- Smith, P. (1997). "High-performance ambulances provide an intelligent response." Mapping Awareness 11(2): 19-21.
- Srinivasan, D., A. C. Liew and C. S. Chang (1994). "A neural network short-term load forecaster." Electric Power Systems Research 28: 227-234.
- Stefánsson, A., Koncar, N., Jones, A. J. (1997). "A Note of the Gamma Test." Neural Computing and Applications 5: 131-133.
- Sutherland, E. H. and D. R. Cressey (1970). Principles of criminology, Lippincott: Philadelphia.
- Tazelaar, J. M. (1989). Neural networks. Byte: 214.
- The-Stationery-Office (1998). Crime and Disorder Act. London, Home Office 1998.
- USDOJ (1997). Spatial analysis of crime (SCAS), US Department of Justice, Criminal Division.
- Van Koppen, P. J. and R. W. J. Jansen (1999). "The time to rob: Variations in time of number of commercial robberies." Journal of Research in Crime and Delinquency 36(1): 7-29.
- Wallis, C. P. and R. Maliphat (1967). "Delinquent areas in the county of London." Delinquent areas in the county of London 7: 250-284.

- Weisburn, D. and J. T. McEwen (1995). Crime mapping and crime prevention. Monsey, New York, Criminal Justice Press.
- Wilson, I. D., Paris, S.D., W. J.A. and D. H. Jenkins (2002). "Residential property price time series forecasting with neural networks." Knowledge-Based-Systems **15**: 335-341.
- Wilson, J. Q. and G. E. Kelling (1982). Broken windows: The Police and Neighbourhood Safety. Atlantic Monthly.
- Worton, B. J. (1987). "A review of models of home range for animal movement." Ecological Modelling **38**: 277-298.
- Worton, B. J. (1989). "Kernel methods for estimating the utilization distribution in home-range studies." Ecology **70**: 164-168.
- Wright, M. S. (1997). Development and trial of a risk assessment toolkit for the UK fire service, Home Office - Fire Research and Development Group.
- Wright, M. S. (1999a). Dwelling risk assessment toolkit, Home Office - Fire Research And Development Group.
- Wright, M. S. (1999b). High occupancy building risk assessment toolkit, Home Office - Fire Research And Development Group.
- Zhang, X. (1994). "Time series analysis and prediction by neural networks." Optimisation Methods and Software **4**: 151-170.
- Zurada, J. M. (1992). Introduction to Artificial Neural Network Systems, West Publishing Company.

GIS Research Group  
University of Glamorgan  
Wales  
United Kingdom  
CF37 1DL

## Crime Mapping Survey

### PART A – BACKGROUND

Constabulary  
Name

i. How many Divisions are there within your Constabulary?

ii. What is the total number of police officers and civilian staff employed by your Constabulary?

iii. What is the approximate population size of the communities your Constabulary serves?

### PART B – ANALYTICAL OPERATIONS

1. What types of crime analysis is the constabulary involved with?

Cross (X) if applicable

☐ Point Pattern analysis

☐ Case Studies

☐ Other (Specify)

☐ Pin Maps

☐ Statistical Reports

☐ Trend analysis

☐ Pattern Detection

2. Is the Constabulary involved in computerised crime mapping

☐ YES

☐ NO (Go to Part H)

2. (b) Are all Divisions within the Constabulary engaged in computerised crime mapping?

☐ YES (Go to Question 3)

☐ NO

2. (c) How many are involved?

3. Where is crime mapping carried out?

Cross (X) if applicable

☐ At Headquarters

☐ Other (Specify)

☐ In a specific Division

☐ In All Divisions

4. How long has the Constabulary been involved in computerised mapping?

**5. Who performs the computerised mapping?***Cross (X) if applicable*

- ☐ Specialised staff  
☐ Police Officers  
☐ Other (*Specify*)

**PART C – SOFTWARE****6. Does the Constabulary use a single type of mapping software?**

- ☐ YES (*Go to Question 7*)  
☐ NO

**6. (b) How many different types of mapping software do you use?**

**7. Is the mapping software a commercially available product?**

- ☐ YES  
☐ NO (*Go to section X*)  
☐ Mixture

**8. What commercial packages are used?***Cross (X) if applicable**Package**Version*

<input type="checkbox"/> ArcInfo	<input type="text"/>
<input type="checkbox"/> ArcView	<input type="text"/>
<input type="checkbox"/> ER Mapper	<input type="text"/>
<input type="checkbox"/> Intergraph	<input type="text"/>
<input type="checkbox"/> MapInfo	<input type="text"/>
<input type="checkbox"/> IDRISI	<input type="text"/>
<input type="checkbox"/> Other ( <i>Specify</i> )	<input type="text"/>

**9. How many terminals does the Constabulary maintain for crime mapping?**

**10. Do you have an intranet?**

- ☐ YES  
☐ NO (*Go to Question 12*)

**11. (b) Is any crime mapping available over the intranet?**

- ☐ YES  
☐ NO (*Go to Question 12*)

**11. (c) Approximately what percentage of staff has direct access to the intranet?***Cross (X) if applicable*

- ☐ < 30%  
☐ 30-50%  
☐ 50-70%  
☐ >70%



**12. Has your Constabulary been involved in customisation of any mapping software?**

- ☐ YES  
☐ NO (Go to Question 13)

**12 (b) Was the customisation carried out in-house?**

- ☐ YES  
☐ NO

**12. (c) What types of customisation have been undertaken?**

*Cross (X) if applicable*

- ☐ Specialised data integration software      ☐ Other (Specify)  
☐ Specialised visualisation techniques  
☐ Specialised querying tools

**13. What types of database management packages does the Constabulary employ?**

*Cross (X) if applicable*

- ☐ Dbase                              ☐ Oracle                              ☐ Other (Specify)  
☐ FoxPro                              ☐ Paradox  
☐ Microsoft Access              ☐ Sybase

**14. Do all Divisions utilise the same database management software?**

- ☐ YES  
☐ NO

**15. Do all Divisions have access to a central Constabulary-wide database(s)?**

- ☐ YES  
☐ NO

**15. (b) Does each Division maintain its own distinct database?**

- ☐ YES  
☐ NO

**16. Does the Constabulary or any of its Divisions utilise a global positioning system (GPS) in any operational context?**

- ☐ YES  
☐ NO (Go to Part D)

**16. (b) How do you utilise GPS technology?**

*Cross (X) if applicable*

- ☐ Car location system  
☐ Call for service identification  
☐ Other (Specify)

**PART D – CRIME ANALYSIS****17. What type(s) of data does the Constabulary map?***Cross (X) if applicable*

- ☐ Command and Control  
☐ Offence data  
☐ Other (*Specify*)

**18. If offence data is mapped, what type(s) are mapped?***Cross (X) if applicable*

- |   |  |   |
|---|--|---|
| <input type="checkbox"/> Arson                    | <input type="checkbox"/> Drug related offences | <input type="checkbox"/> Sexual offences    |
| <input type="checkbox"/> Assault                  | <input type="checkbox"/> Domestic Violence     | <input type="checkbox"/> Disorderly conduct |
| <input type="checkbox"/> Burglary                 | <input type="checkbox"/> Firearms offences     | <input type="checkbox"/> Traffic offences   |
| <input type="checkbox"/> Other ( <i>Specify</i> ) |  |   |

**19. What types of mapping analysis are performed***Cross (X) if applicable*

- ☐ Automatic point maps                      ☐ Other (*Specify*)  
☐ Hot Spot analysis  
☐ Temporal analysis

**20. How often is crime analysis conducted?***Cross (X) if applicable*

- ☐ Daily                      ☐ Monthly  
☐ Weekly                      ☐ Other (*Specify*)  
☐ Fortnightly

**21. How are the results from crime mapping analysis used?***Cross (X) if applicable*

- ☐ Inform police officers                      ☐ Other (*Specify*)  
☐ Evaluation of policies  
☐ Identification for resource allocation

**22. Are you involved in any cross-Constabulary co-ordination in terms of crime mapping?**

- ☐ YES  
☐ NO

**23. Rate the usefulness of crime analysis in the context of achieving the Constabulary mission and objectives. (Circle an appropriate number 1= Not Useful 5 = Extremely Useful)**

	1 Not Useful	2	3	4	5 Extremely Useful	
--	-----------------	---	---	---	-----------------------	--

**PART E – CONTEXTUAL DATA****24. What street map data source(s) are used for crime mapping?***Cross (X) if applicable*

- ☐ Commercially available (*Specify*)  
☐ Developed in-house (*Specify*)  
☐ Other (*Specify*)

**25. Are any additional spatial data sets used in conjunction for crime mapping?***Cross (X) if applicable*

- |   |   |
|---|---|
| <input type="checkbox"/> Aerial Photographs | <input type="checkbox"/> Parks information        |
| <input type="checkbox"/> CCTV coverage      | <input type="checkbox"/> Satellite Imagery        |
| <input type="checkbox"/> Census data        | <input type="checkbox"/> Urban Priority Areas     |
| <input type="checkbox"/> Land Use maps      | <input type="checkbox"/> Other ( <i>Specify</i> ) |

**26. Rate the usefulness of contextual data in your crime analysis activities (Circle an appropriate number 1= Not Useful 5 = Extremely Useful)**

	1 Not Useful	2	3	4	5 Extremely Useful	
--	-----------------	---	---	---	-----------------------	--

**PART F – OVERALL****27. Rate each of the following statements according to the extent to which they have a negative impact upon the Constabulary and its ability to use crime mapping effectively. (Circle an appropriate number 1 = Unlikely 5 = Serious problem)**

Limited computer resources	1 No Problem	2	3	4	5 Serious Problem
Limited financial resources	1 No Problem	2	3	4	5 Serious Problem
Limited time	1 No Problem	2	3	4	5 Serious Problem
Limited training opportunities	1 No Problem	2	3	4	5 Serious Problem
Limited working knowledge of how mapping is used in the field	1 No Problem	2	3	4	5 Serious Problem
Limited interest from administration	1 No Problem	2	3	4	5 Serious Problem
Limited interest from support staff	1 No Problem	2	3	4	5 Serious Problem
Difficulties with computer software	1 No Problem	2	3	4	5 Serious Problem
Other ( <i>Specify</i> )	1 No Problem	2	3	4	5 Serious Problem
	1 No Problem	2	3	4	5 Serious Problem
	1 No Problem	2	3	4	5 Serious Problem

**28. Rate each of the following statements concerning an overall perception of crime mapping and its support (Circle an appropriate number 1 = Inaccurate 5 = Very Accurate)**

Leadership financially supports crime mapping	1 Inaccurate	2	3	4	5 Very Accurate
Crime mapping is a valuable tool	1 Inaccurate	2	3	4	5 Very Accurate
Investment of resources into crime mapping should continue	1 Inaccurate	2	3	4	5 Very Accurate

**PART G – FURTHER COMMENTS**

☐ I would like to receive a summary of the results

**Thank you for taking the time to complete this survey. Please return the completed survey in the pre-addressed envelope provided.**

**PART H – CONSTABULARIES NOT ENGAGED IN COMPUTERISED CRIME MAPPING****29. Are there future plans to incorporate any crime mapping facilities?**

- ☐ YES  
☐ NO (Go to Question 30)

**29. (b) What is time scale planned to incorporate mapping technology**

*Cross (X) if applicable*

- ☐ Within the next year  
☐ 1-2 Years  
☐ 3-5 Years  
☐ Other (*Specify*)

**29. (c) What areas are planned for computerised mapping incorporation?**

*Cross (X) if applicable*

- ☐ Command and Control  
☐ Crime Analysis  
☐ Resource allocation  
☐ Other (*Specify*)

**30. Do you maintain a computerised archive of crime data?**

- ☐ YES  
☐ NO (Go to Question 32)

**30. (b) What types of database management packages does the Constabulary employ?**

*Cross (X) if applicable*

- |   |                                  |   |
|---|----------------------------------|---|
| <input type="checkbox"/> Dbase            | <input type="checkbox"/> Oracle  | <input type="checkbox"/> Other ( <i>Specify</i> ) |
| <input type="checkbox"/> FoxPro           | <input type="checkbox"/> Paradox |   |
| <input type="checkbox"/> Microsoft Access | <input type="checkbox"/> Sybase  |   |

**30. (c) Do all Divisions utilise the same database management software?**

- ☐ YES  
☐ NO

**30 (d) Do all Divisions have access to a central Constabulary-wide database(s)?**

- ☐ YES (Go to Question 32)  
☐ NO

**30. (e) Does each Division maintain its own distinct database?**

- ☐ YES  
☐ NO

**31. Do you currently geocode any of your crime data?**

- ☐ YES  
☐ NO (Go to Question 32)

**31.(a) What period of time have you geocoded your crime data?**

*Cross (X) if applicable*

- ☐ Under 12 months  
☐ 1-2 Years  
☐ 3-5 Years  
☐ Other (*Specify*)

**32. Rate the perceived usefulness of crime mapping in the context of your Constabulary (Circle an appropriate number 1= Not Useful 5 = Extremely Useful)**

	1 Not Useful	2	3	4	5 Extremely Useful	
--	-----------------	---	---	---	-----------------------	--

**33. Rate each of the following statements according to the extent to which they have a negative impact upon the Constabulary and its ability to use crime mapping effectively. (Circle an appropriate number 1 = Unlikely 5 = Serious problem)**

Limited computer resources	1 No Problem	2	3	4	5 Serious Problem
Limited financial resources	1 No Problem	2	3	4	5 Serious Problem
Limited time	1 No Problem	2	3	4	5 Serious Problem
Limited training opportunities	1 No Problem	2	3	4	5 Serious Problem
Limited working knowledge of how mapping is used in the field	1 No Problem	2	3	4	5 Serious Problem
Limited interest from administration	1 No Problem	2	3	4	5 Serious Problem
Limited interest from support staff	1 No Problem	2	3	4	5 Serious Problem
Difficulties with computer software	1 No Problem	2	3	4	5 Serious Problem
Other ( <i>Specify</i> )	1 No Problem	2	3	4	5 Serious Problem
	1 No Problem	2	3	4	5 Serious Problem
	1 No Problem	2	3	4	5 Serious Problem

**PART I – FURTHER COMMENTS**

☐ I would like to receive a summary of the results

**Thank you for taking the time to complete this survey. Please return the completed survey in the pre-addressed envelope provided.**

**Police Survey**

The questionnaire was initially sent out on 7<sup>th</sup> November 2000 to all 52 UK Constabularies. The response following the first mailing was 29%, followed by 19% from the second mailing sent on 4<sup>th</sup> December 2000. A final mailing was sent on 11<sup>th</sup> January 2001 yielding an additional 4%, producing an overall 52% response to the questionnaire.

**Key Facts**

<b>BACKGROUND</b>				
% of responding Constabularies			52%	
Average number of divisions			6	
Average number of employees (civilian & police)			3,253	
Average population coverage			922,570	
<b><u>MAPPING CONSTABULARIES</u></b>				
<b>MAPPING and SOFTWARE</b>				
% engaged in mapping			85%	
Most popular mapping software			SerWorld (48%) MapInfo (35%) PAFEC (22%)	
Most popular database software			Oracle (70%) Microsoft Access (61%) SQL Server (35%) Ingress (26%)	
Average number of terminals dedicated to mapping			56	
% involved in customisation of mapping software			74%	
% integrating GPS technology			30%	
<b>CRIME ANALYSIS</b>				
Use of mapping			Offence data (96%) Command & Control (74%) Road Traffic Accident data (30%)	
Most popular type of crime analysis			Hot-spot analysis (83%)	
Regularity of crime analysis			Daily (78%), Weekly(13%), Monthly(13%)	
Most popular use of crime analysis			Inform Police Officers (87%) Evaluate policies (52%) Identify resource requirements (65%)	
% involved in cross-constabulary mapping			13%	
<b>CONTEXTUAL DATA</b>				
Most popular types of contextual data			Census data (26%) CCTV Coverage (13%) Aerial photographs (13%)	
<b>OVERALL</b>				
Perceived usefulness of crime mapping 1(Not Useful) – 5(Extremely Useful)				
1	2	3	4	5
5%	11%	37%	26%	21%
<b><u>NON-MAPPING CONSTABULARIES</u></b>				
<b>FUTURE PLANS</b>				
% planning to acquire mapping software			100%	
Timescale for acquisition			50% within 12 months 50% within 1-2 years	
Planned uses for mapping			Command and Control (100%) Crime Analysis (100%) Resource Allocation (25%)	
Perceived usefulness of crime mapping 1(Not Useful) – 5(Extremely Useful)				
1	2	3	4	5
0%	0%	0%	0%	100%



**Fire Survey**

The questionnaire was initially sent out on 2<sup>nd</sup> May 2001 to each of the 58 UK Fire brigades. The response following the first mailing was 28%, followed by 26% from the second mailing sent on 8<sup>th</sup> June 2001 producing an overall 54% response to the questionnaire.

**Key Facts**

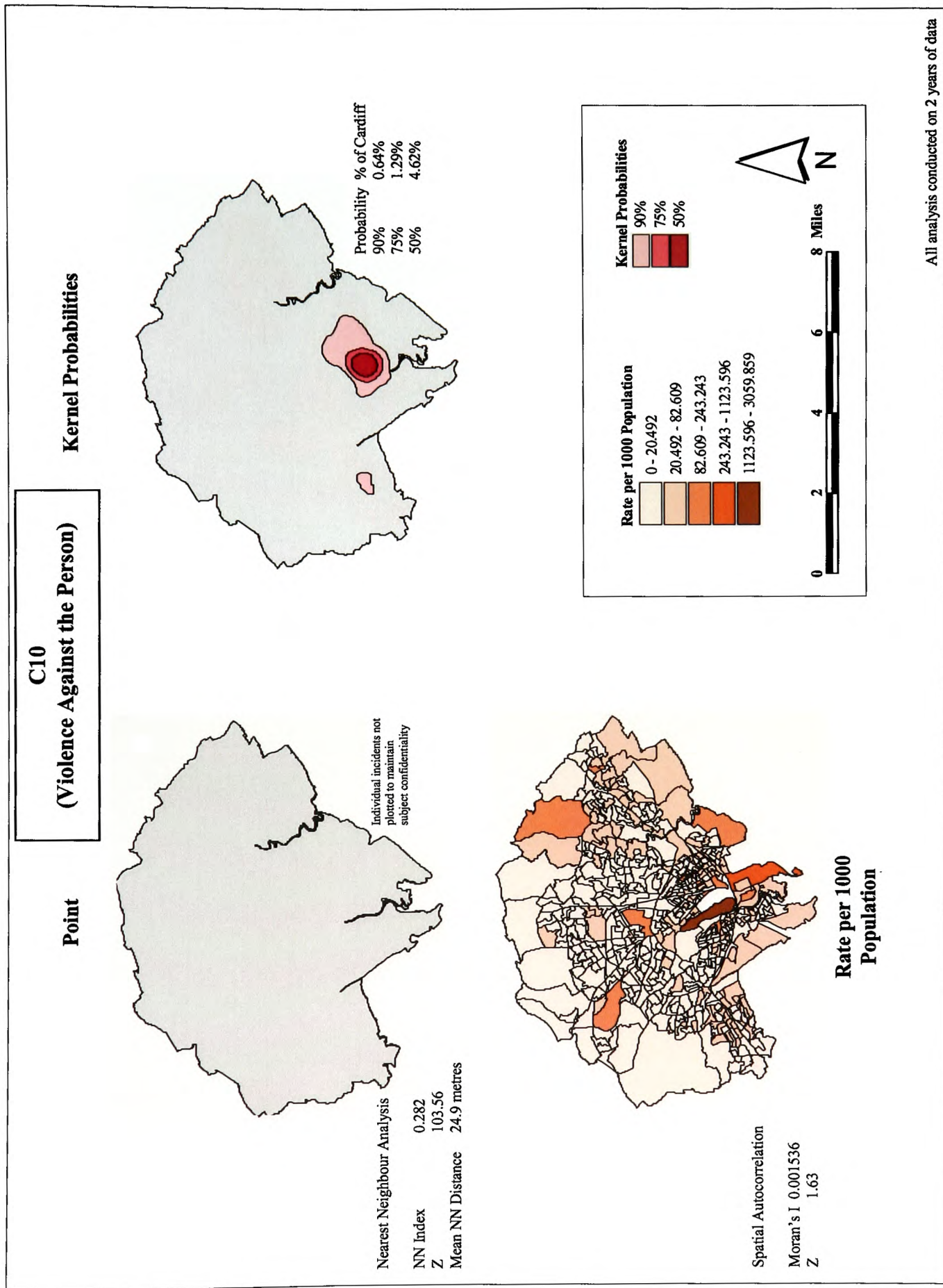
<b>BACKGROUND</b>				
% of responding Brigades			54%	
Average number of divisions			3	
Average number of employees (civilian & fire officers)			1,372	
Average population coverage			1,165,478	
<b><u>MAPPING BRIGADES</u></b>				
<b>MAPPING and SOFTWARE</b>				
% engaged in mapping			94%	
Most popular mapping software			MapInfo (41%) SerWorld (31%) Other (72%)	
Most popular database software			MS Access (91%) Oracle (45%) MS SQL Server (41%)	
Average number of terminals dedicated to mapping			12	
% involved in customisation of mapping software			62%	
% integrating GPS technology			24%	
<b>INCIDENT ANALYSIS</b>				
Use of mapping			Command and Control (69%) Incident Analysis (90%) Other (38%)	
Most popular type of incident analysis			Hot Spot analysis (69%)	
Regularity of incident analysis			Daily (17%), Weekly (3%), Monthly (28%), Other - Ad Hoc (41%)	
Most popular use of incident analysis			Inform fire officers (69%) Evaluation of policies (48%) Identification for resource allocation (52%)	
% involved in cross-brigade mapping			7%	
<b>CONTEXTUAL DATA</b>				
Most popular types of contextual data			Aerial photography (14%) CCTV coverage (14%) Land use mapping (14%)	
<b>OVERALL</b>				
Perceived usefulness of computerised mapping 1(Not Useful) – 5(Extremely Useful)				
1	2	3	4	5
4%	4%	27%	31%	34%
<b><u>NON-MAPPING BRIGADES</u></b>				
<b>FUTURE PLANS</b>				
% planning to acquire mapping software			100%	
Timescale for acquisition			Within the next year (50%) 1-2 years (50%)	
Planned uses for mapping			Incident Analysis (100%) Command and Control (50%) Resource Allocation (50%)	
Perceived usefulness of computerised mapping 1(Not Useful) – 5(Extremely Useful)				
1	2	3	4	5
0%	0%	0%	100%	0%

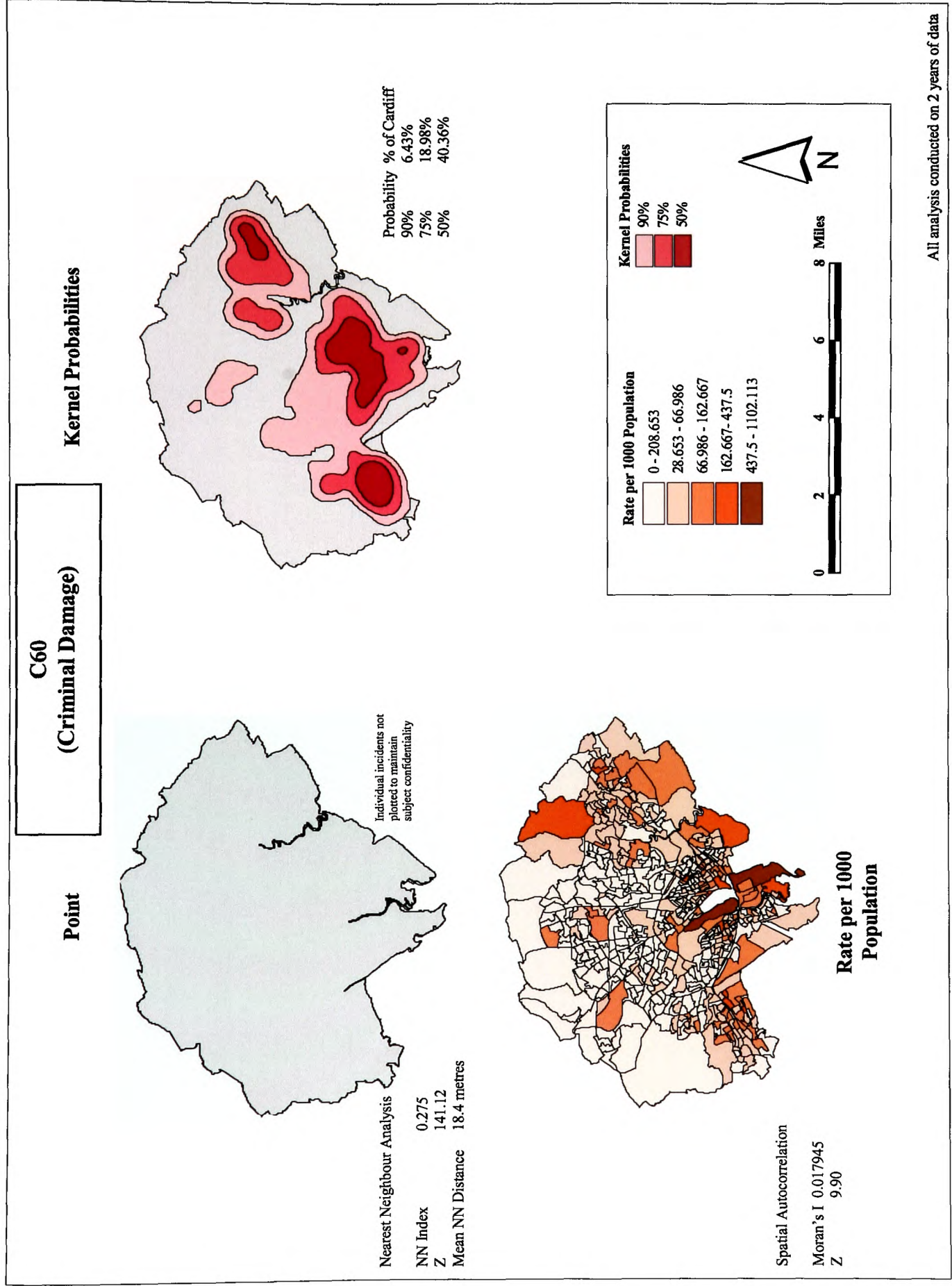
**Ambulance Survey**

The questionnaire was initially sent out on 4<sup>th</sup> May 2001 to all UK Ambulance Trusts and Services. The response following the first mailing was 31%, followed by 23% from the second mailing sent on 8<sup>th</sup> June 2001, producing an overall 54% response to the questionnaire.

**Key Facts**

<b>BACKGROUND</b>				
% Of responding Trusts		54%		
Average number of stations		26		
Average number of employees		925		
Average population coverage		1,774,066		
<b>MAPPING and SOFTWARE</b>				
% engaged in mapping		100%		
Most popular mapping software		PAFEC / SerWorld / Blue8 (37%) MapInfo (26%) ArcView (11%)		
Most popular database software		MS Access (79%) Oracle (47%) MS SQL Server (42%)		
Average number of terminals dedicated to mapping		17		
% involved in customisation of mapping software		56%		
% integrating GPS technology		68%		
<b>INCIDENT ANALYSIS</b>				
Use of mapping		Command and Control (100%) Incident Analysis (74%) Other (21%)		
Most popular types of incident analysis		Hot Spot analysis (79%) Automatic point maps (47%)		
Regularity of incident analysis		Daily (58%), Weekly (11%), Monthly (26%), Other - Ad Hoc (21%)		
Most popular use of incident analysis		Inform Ambulance staff (53%) Evaluation of policies (58%) Identification for resource allocation (100%)		
% involved in cross-Trust mapping		26%		
<b>CONTEXTUAL DATA</b>				
Most popular types of contextual data		Census Data (16%) Aerial photography (5%)		
<b>OVERALL</b>				
Perceived usefulness of computerised mapping 1(Not Useful) – 5(Extremely Useful)				
1	2	3	4	5
0%	0%	11%	50%	39%

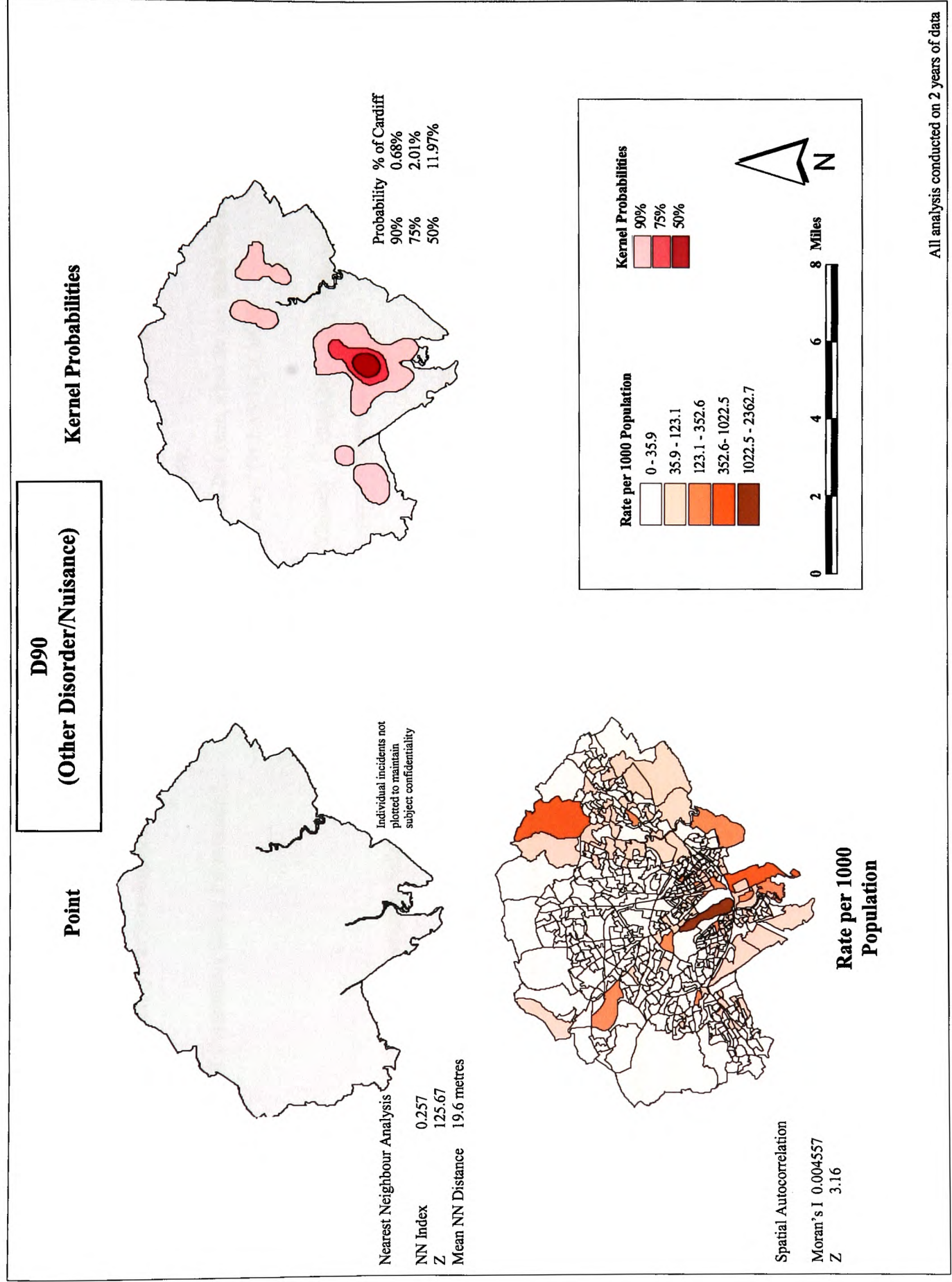




All analysis conducted on 2 years of data







All analysis conducted on 2 years of data

*Gweirthio gyda 'n gilydd i greu dinas fwy diogel*

*Working together for a safer city*



## Community Safety Consultation

The Cardiff Community Safety Partnership aims to reduce crime, disorder and their economic costs in the City and County of Cardiff in a cost-effective and socially equitable way.

1. Are you aware of the Cardiff Community Safety Partnership?

(PLEASE TICK BOX)

Yes ☐ No ☐

2. If so, how did you hear of the Cardiff Community Safety Partnership?

(PLEASE CIRCLE)

Newspaper      Radio      Television  
Local Neighbourhood Watch      Local Crime Prevention Panel  
Other.....

3. Please rank in order of importance the issues you feel the Cardiff Community

Safety Partnership should focus upon in your area?

(Please number from 1, the most important to 10, the least important)

Car crime ☐ Burglary ☐  
Young people ☐ Violent crime ☐  
Racial attacks ☐ Attacks on gay men and lesbians ☐  
Drug abuse ☐ Domestic violence ☐  
Noise nuisance ☐ Neighbourhood nuisance ☐

4. Over the last year, what do you think has happened to the level of crime in your area? (PLEASE TICK BOX)

☐ Greatly ☐ Slightly ☐ Unchanged ☐ Slightly ☐ Greatly  
Increased Increased Decreased Decreased

5. How high/low would you say the rate of the following crimes are:

(PLEASE TICK BOX)

	In Your Neighbourhood				In Cardiff as a whole			
	Very High	Fairly High	Fairly Low	Very Low	Very High	Fairly High	Fairly Low	Very Low
E.g. Noise Nuisance								
Home broken into								
Mugging on the street								
Vehicle stolen								
Sexual Assault or Rape								
Theft from vehicle								

## 5. Continued

	In Your Neighbourhood					In Cardiff as a whole				
	Very High	Fairly High	Fairly Low	Very Low		Very High	Fairly High	Fairly Low	Very Low	
Physical attack										
Racist attack										
Attacks against gay men and lesbians										
Violence against partner or spouse										
Pestered or insulted in public										
Neighbourhood nuisance										

## 7. How safe do you feel walking alone in the City Centre?

(PLEASE TICK)

	Very Safe	Fairly Safe	Not Very Safe	Not Safe At All
Morning (7am – 12pm)				
Afternoon (12pm – 5pm)				
Evening (5pm – 10pm)				
Late Night (10pm – 3am)				

## 6. How safe do you feel walking alone in your neighbourhood?

(PLEASE TICK)

	Very Safe	Fairly Safe	Not Very Safe	Not Safe At All
Morning (7am – 12pm)				
Afternoon (12pm – 5pm)				
Evening (5pm – 10pm)				
Late Night (10pm – 3am)				

## 8. How safe do you feel walking alone in Cardiff Bay?

(PLEASE TICK)

	Very Safe	Fairly Safe	Not Very Safe	Not Safe At All
Morning (7am – 12pm)				
Afternoon (12pm – 5pm)				
Evening (5pm – 10pm)				
Late Night (10pm – 3am)				



**9. Have you been a victim of any of the following crimes or incidents in the last 12 months?**  
(PLEASE TICK)

	Victim	
	Yes	No
Home broken into		
Violence against partner/spouse		
Neighbourhood nuisance		

**10. Have you been a victim or witness of any of the following crimes or incidents in the last 12 months?**  
(PLEASE TICK)

	Victim			Witness	
	Yes	No		Yes	No
<b>Mugging in the street</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				
<b>Vehicle stolen</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				

	Victim			Witness	
	Yes	No		Yes	No
<b>Racist attack</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				
<b>Attacks against gay men and/or lesbians</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				
<b>Theft from a vehicle</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				
<b>Physical attack</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				
<b>Pestered or insulted in public</b>	in your neighbourhood				
	elsewhere in Cardiff				
	elsewhere				

11. Are you      Female ☐      Male ☐

12. How old are you?

.....

13. Now thinking about your current employment situation, are you?  
(PLEASE TICK)

Employed/self employed full time	<input type="checkbox"/>	Retired	<input type="checkbox"/>
Employed/self employed part time	<input type="checkbox"/>	Unemployed	<input type="checkbox"/>
In full time education	<input type="checkbox"/>	Keeping home	<input type="checkbox"/>

14. Which of the following best describes your ethnic origin?  
(PLEASE TICK)

Black Caribbean	<input type="checkbox"/>	Black African	<input type="checkbox"/>
Black other	<input type="checkbox"/>	Indian	<input type="checkbox"/>
Pakistani	<input type="checkbox"/>	Bangladeshi	<input type="checkbox"/>
Chinese	<input type="checkbox"/>	Asian	<input type="checkbox"/>
White	<input type="checkbox"/>	Traveller	<input type="checkbox"/>
Other .....			

15. Which of the following best describes your household?  
(PLEASE TICK)

Couple living with children	<input type="checkbox"/>	Living alone	<input type="checkbox"/>
Couple/living with partner	<input type="checkbox"/>	Single parent	<input type="checkbox"/>
Sharing with others (e.g. student)	<input type="checkbox"/>		

16. What type of accommodation do you live in?  
(PLEASE TICK)

Ground floor flat	<input type="checkbox"/>	Flat/maisonette	<input type="checkbox"/>
House	<input type="checkbox"/>	Mobile home	<input type="checkbox"/>
Other .....			

17. Over the last 12 months do you consider your health to have been  
(PLEASE CIRCLE)

Very good	Fairly good	Fairly poor	Very poor
-----------	-------------	-------------	-----------

18. Please state your street name and/or postcode

..... CF.....

19. If you would like to comment on this questionnaire or community safety,  
we would appreciate your remarks

.....

.....

.....

.....

.....

*Thank you for helping us reduce crime and disorder in Cardiff and  
completing this consultation*

## **PUBLISHED PAPERS**

### **PAPER 1**

**Corcoran, J. and J. A. Ware (2002)**

“Helping with enquiries.” GeoConnexion. 1(3): 36-40.

---

### HELPING WITH ENQUIRIES

J. Corcoran. School of Computing, University of Glamorgan, Pontypridd, Wales, UK

J.A. Ware. School of Computing, University of Glamorgan, Pontypridd, Wales, UK

The mapping of crime has a long history as a tool for understanding its spatial distributions. It can be traced back as far as the 19th century in France, where hand drawn maps were utilised to visualise and analyse crime information. In time this so called "*cartographic school of criminology*" spread to Britain although it dissipated soon after, the traditional focus on crime redefined to sociological and psychological based enquiries.

Since the late 1960s the advent of computing technology and more latterly the development of Geographical Information System (GIS) technologies, have undoubtedly fuelled the proliferation of a range of crime mapping and analysis systems. The flexibility, powerful visualisation and integral capacities that these systems offer have thus assisted the restatement in importance of a geographically orientated approach to crime analysis.

The use of computer based crime mapping technologies by the Police is rapidly becoming a vital prerequisite in understanding incident distributions and assisting in both the identification / allocation of resources and production / evaluation of policing strategies. The ability to efficiently generate simple maps to depict crime location and densities can be used directly to inform police officers and policing strategies, therefore maximising officer and strategy effectiveness and potential. A recent Home Office report underlined the importance of geographic data for analysis and interpretation of crime at the local level. It states that through its use Police Forces will be able to conduct a more detailed analysis of their data to identify hot-spots and in the allocation of resources, in addition to identifying incidents occurring across the various Police boards.

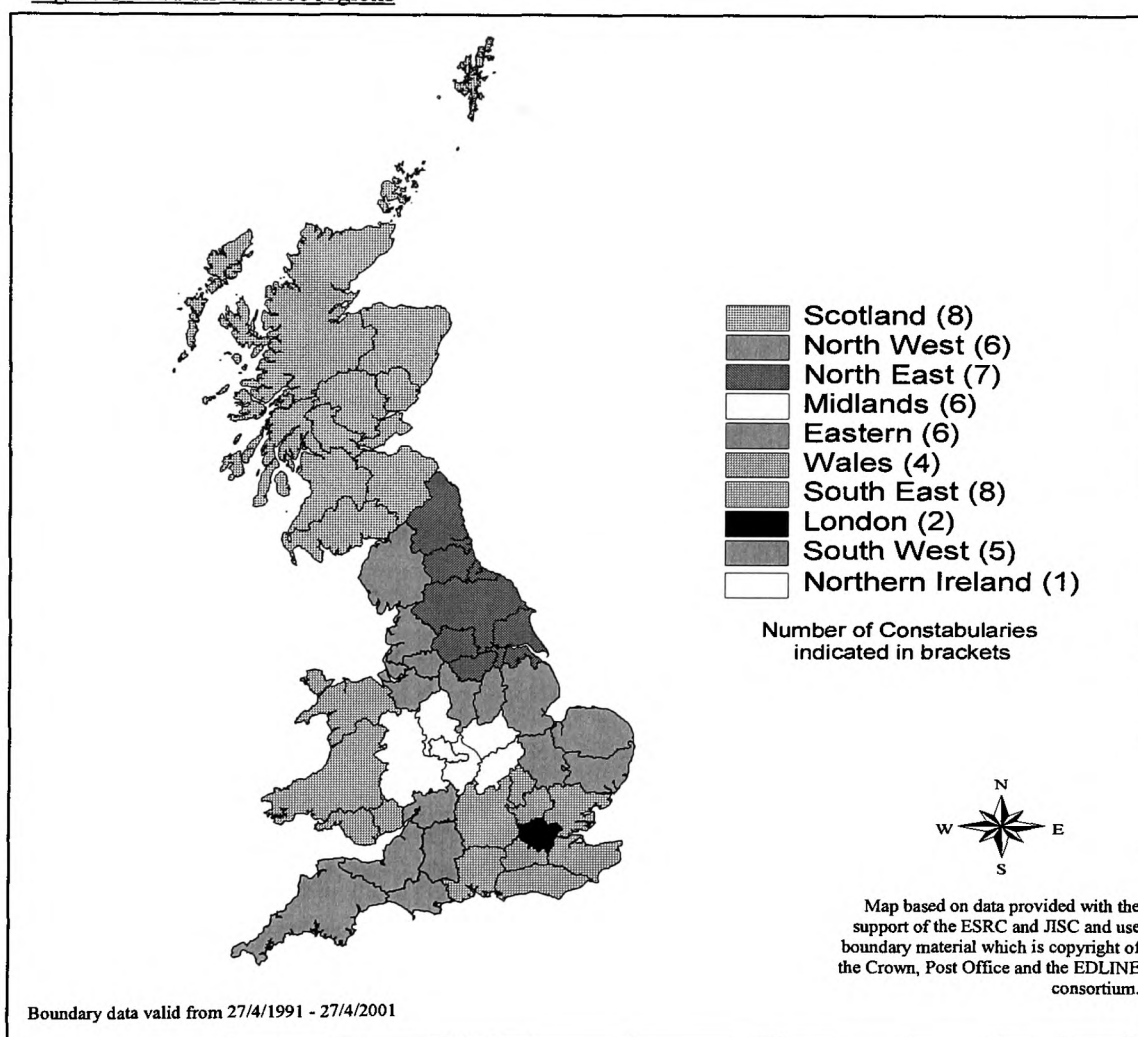
Previous to this survey there was no tangible evidence portraying the degree to which computer based crime mapping has been adopted, integrated and utilised by UK Police Forces despite Governmental

promotion of such technology. This survey aims to provide a current quantification of its use, in association with the problems and difficulties encountered.

### UK Police Force structure

The UK is geographically segregated into 10 generic areas consisting of 52 Constabulary regions, each of which is further sub-divided into a series of territorial divisions under Constabulary jurisdiction (figure 1).

Figure 1 UK Police Force regions



This survey targeted the Constabulary level on the basis that this would provide a comprehensive overview of all crime mapping activities at both local and regional levels of jurisdiction and administration, in the context of national policy. The survey tool was thus designed to pose

Constabulary-wide questions in relation to generic activities, strategies and future policy in addition to previous, current and perceived future engagement with computer based crime mapping.

### **Mapping or no Mapping?**

Overall the survey generated a **52%** response rate equating to **27** Constabularies across the UK. Unfortunately total UK coverage was not achieved, indication from some Forces not participating was either lack of time or a reluctance to disclose certain intelligence information to questions posed by the survey tool.

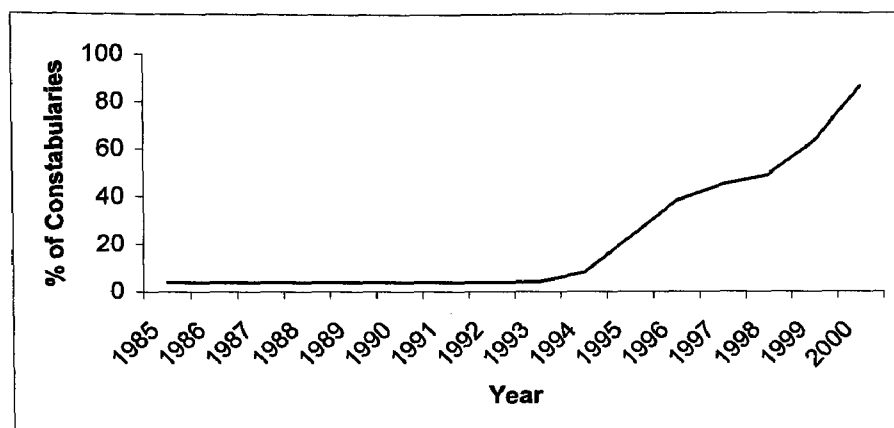
Of the respondents, **85%** are currently engaged in some form of computerised crime mapping. The US survey indicated a figure of 13%<sup>1</sup> of departments engaged in computer based crime mapping, contrasting to a figure of **48%** indicated by this survey at the same point in time (1998 figure). The disparity between the US and the UK usage represents the department level at which the US survey was targeted where the majority represented agencies employing less than 100 officers. A figure of 36% was reported for those departments employing in excess of 100 officers, providing a more comparable sample to that represented in the UK survey.

Acquisition of mapping software was shown to be a relatively recent occurrence with a rapid increase in use within the past 5 years. Of the **85%**, a total of **44%** acquired their computer based mapping technology since 1998 and **91%** since 1994 (figure 2). For those Forces not currently engaged in computerised crime mapping all intend to invest in such technology, **50%** of which plan to do so within the next 12 months, the remainder within a 1-2 year period.

For those Forces engaged in computerised crime mapping activities **74%** have implemented a Constabulary wide mapping facility incorporating all divisions, **22%** deploying a single divisional or head-quarters installation. The remaining **4%** represent either Forces in their early stages of the technology deployment or those wishing to centralise all computerised crime mapping activities.

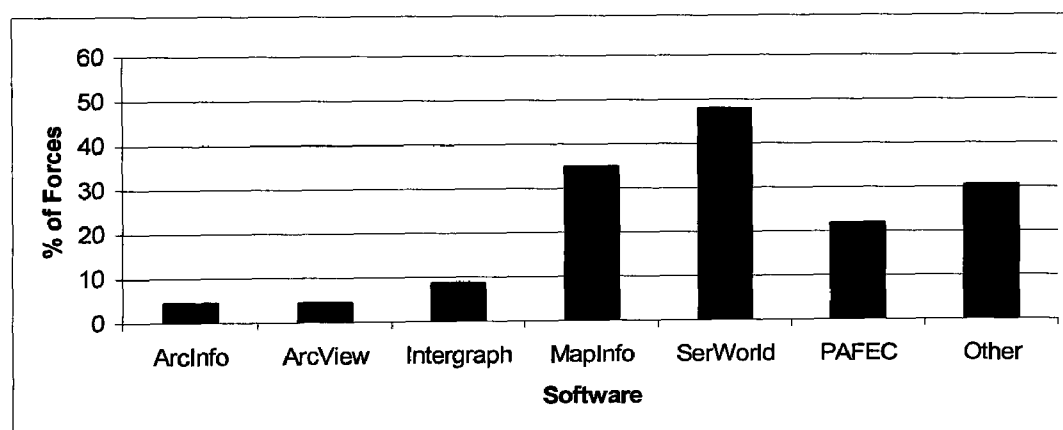
---

<sup>1</sup> Statistics taken from (Mamalian, 1999) pg.1



**Figure 2** Timeline of computerised crime mapping technology acquisition.

Selection of mapping software was shown to be somewhat divided with **48%** of Forces utilising two or more packages, SERworld and MapInfo® proving the most popular, a relatively large number opting for a miscellaneous of packages (figure 3). For Forces utilising more than a single software provider (**43%**) there was no specific combination proving more popular.



**Figure 3** Types of mapping software used

A major cost consideration in the selection of mapping software is its ability to interface with existing data sources demanding both minimal modification to hardware / software and alteration to existing working practices. One major consideration is the database (or in many cases databases) storing the information to be mapped, and their compatibility with the mapping software. The survey revealed that the most popular database was Oracle (**70%**), followed by Microsoft Access (**61%**) and Microsoft SQL Server (**35%**). The majority of Forces (**70%**) possess multiple types of database software, the most common configuration, a corporate Oracle database and instances of Microsoft Access for local

access and data manipulation, accounting for **56%**. A similar database arrangement was also reported by all those Forces not currently engaged in computer based crime mapping.

The number of terminals dedicated for crime mapping activities varied enormously from Force to Force ranging from single figures to over **500**, the average being **56**. This figure appeared to be influenced to some degree by a series of key variables, including length of time the Force has been engaged in mapping (the greater the time the greater the number of terminals) and the presence of technical resources, in particular the availability of mapping via Intranet facilities.

### **Dissemination**

Successful integration and utilisation of any mapping software within an operational policing context is reliant upon both technical and strategic knowledge in addition to a comprehensive plan for its exploitation on a daily, weekly, monthly basis. The survey indicated that **91%** of Forces found it necessary to employ specialised staff to conduct computerised crime mapping, **43%** employ both specialised staff and police officers and **9%** of Forces utilise only police officers. Key to the effective use of computerised crime mapping is efficient dissemination of its outputs that may be in the form of point or hot-spot maps to resource managers and patrol officers. One potential mechanism for achieving this is through the use of a secure Intranet that offers an efficient, low maintenance and economical method for dissemination to all necessary parties. The survey indicated that all Forces possess an Intranet (**43%** of Forces reporting that in excess of **70%** of their staff have direct access to such facilities) although only **35%** utilise this facility to disseminate output from their mapping activities.

The regularity with which analysis of crime data is conducted can potentially impact upon its effectiveness as a tactical tool for policing and its ability to convey accurate, timely intelligence concerning the status of a region. This was reflected in the survey whereby **78%** of all Forces conducted crime analysis daily, **13%** weekly and **9%** on a monthly basis. The currency of spatial data and cartographic output is of particular importance for tasks such as informing police officers (of which **87%** Forces do so) and to a greater degree for identification of resource requirements such as command and control (of which **65%** Forces are involved), where response within relatively short



timescales is essential. The demand for real, or near real-time, mapping information is reduced somewhat when applied to other areas of policing (such as evaluation of policies and strategies, as carried out by 52% of Forces reflecting, to some extent, the 9% of Forces generating less regular monthly mapping output and spatial analyses).

### **Customisation**

Effective integration and application of mapping software has in certain cases necessitated customisation of the software. The survey suggested that 74% of Forces found it necessary to customise their standard mapping software in order to integrate with existing databases (39%), provide specialised visualisation (4%) or querying tools (43%). Of the 74% of Forces that customised their software 76% resourced the development in-house, the remainder presumably utilising external contractors to complete the development task.

### **Crime Mapping and Analysis techniques**

Geocoding of police data has allowed the visualisation of crimes in a spatial context, prior to which analysis was restricted to text-based searches and pushpins in wall maps. The ability to invoke analytical operations and potentially produce real-time geographic output in the form of density or hot-spot maps is a significant advantage for both operational and strategic policing strategies, *accurate and up-to-date geographic information is becoming a key element of modern policing.*

Crime mapping activities can essentially be divided into 4 categories, where all map offence data, 74% of Forces map command and control information, 30% map road traffic accident data and 17% involved in miscellaneous mapping endeavours. Utilisation and interrogation of this spatio temporal information is through the use of crime analysis, the survey identifying 3 key techniques:

- **Automatic pin maps**

The ability to generate simple point-based mapping depicting individual locations of crimes is a powerful visual tool, which potentially is easily updated and refreshed with new information. Output (dependent upon functionality of the mapping software) can be efficiently tailored according to purpose; for example mapping can be produced to represent a specific type or types of crime, crimes

occurring within a specified distance of a certain type of building or for a specific snapshot in time. Automatic pin-maps therefore offer both a dynamic and highly reusable format by which core information can be queried, aggregated and presented in a variety of styles according to purpose, its popularity as a policing tool reflected by 57% of Forces utilising this technique.

- Hot-spot mapping

Hot-spot mapping compared to automatic-pin map technique represents a greater degree of data processing, although arguably results in the production of a cleaner portrayal of crime volume. The aggregation of point data into a single colour-coded density surface distinguishing areas of high, medium and low crime levels provides a simple and instant visual depiction of a region's most and least vulnerable areas. Its popularity was clearly demonstrated by the survey, 83% of Forces utilising hot-spotting techniques.

- Temporal analysis

Temporal analysis also represents a relatively popular technique, 48% of Forces conducting some form of temporal interrogation of their data. The ability to isolate and identify specific snapshots in time, such as a specific day, date and time then create spatial output representing crime at that instance in time can provide valuable insight into the temporal effects on relative volumes and distribution of crime.

- Other Techniques

A range of other techniques were reported in the survey accounting for 17% of Forces, including predictive and profiling procedures. Although both procedures currently represent a relatively limited exploitation, it is likely as techniques become more established and embedded within existing practices they will receive wider acclaim.

The effectiveness of mapping in terms of its ability to reveal patterns and provide valuable insight into the manifestation of crime and criminal activity is governed by a range of factors. One such variable is the spatial coverage of the input data, where identification of patterns at a local level may preclude patterns evident at district and regional scales. Data coverage and aggregation is therefore key to

enabling mapping and analysis at all scales. One limitation in terms of identifying patterns and trends at broader scales are the territorial limits of the Constabulary within which it operates. To overcome these administrative constraints, cross-Constabulary co-ordination is required, whereby a collaborative mapping agenda is called for, increasing the spatial coverage of map data utilised for pattern and trend detection. The survey, however, revealed that this type of arrangement is relatively limited, 13% of Forces engaged in such activities, the probable reason being a combination of administrative and technical resources required to both initialise and maintain such an operation. It is likely, however, as crime mapping technologies and practices become more established that this figure will increase, working towards a decentralised national collaborative network for mapping and crime analysis.

### **Contextual data**

A fundamental function of any GIS is its ability to offer integration of formerly disparate data sets, enabling visualisation in association with one another thus assimilating a broader picture. The value of this functionality thus lies in its potential to offer valuable insight into the manifestation and patterning of crime and criminal activity through aggregation of various data sets. The ability to visualise a crime locality in relation to land use and physical structures for example is potentially an extremely powerful and valuable policing tool. It was shown that all Forces utilise some form of street map data, an assortment of Ordnance Survey (OS) products (1:1,250, 1:2,500 and 1:10,000 raster tiles, AddressPoint® and OSCAR® road centreline data), Automobile Association (AA) map data and a minority (4%) involved with in-house development of their own street map base data.

In addition to street map data, other contextual data sets such as aerial photographs and census related information can provide valuable insight for the crime analyst. However, the survey showed limited use of such information, aerial photographs (13%), CCTV coverage (13%), census data (26%) and land use mapping (4%). Its overall use was reflected on a scale of 1(Not useful) to 5(Extremely useful), with only 17% rating contextual data as a 5, 17% (4), 30% (3), 9% (2) and 4% (1) believing contextual data not to be of particular importance in crime mapping and analysis activities. It is likely that the cost of acquisition, integration and maintenance of these various data sets currently outweigh their perceived and tangible benefits for the majority of Forces.

**Global Positioning Systems (GPS)**

Location in the form of a spatial reference, is key to mapping applications, the ability to identify a specific locality as a resource or incident of a crime is of prime importance. Integration of GPS technology offers both powerful visual and analytical capabilities through the provision of a calculated and potentially accurate (dependant upon equipment, pre and post-processing techniques and environmental/physical characteristics) spatial reference in the form of an X Y co-ordinate, which can be targeted at identification of resources in the field and/or identification of crime localities.

The use of GPS technology within a police environment is not new, a range of systems have been trailed and implemented both in the United States and United Kingdom. This survey, however, highlighted that its use in an operational context is not widespread, limited to a relatively small number of Forces (30%) where it is utilised for a combination of resource identification/allocation and for routing advice to a specified incident given information on road conditions. It is likely that this figure is set to increase in subsequent years, especially with the fall in cost of GPS technology coupled with the decision to remove the military-induced Selective Availability, therefore as of the 1<sup>st</sup> May 2000 typical commercial GPS position accuracy is approximately 20 metres (opposed to 100 meters prior to this date).

**Perceived usefulness of mapping**

The survey revealed that on a scale of 1(not useful) to 5(extremely useful) 35% of Forces rated the application of computer based mapping technologies (5), 35% (4), 26%(3), 4%(2) and 0%(1). For those Forces not currently engaged in computer based mapping activities all perceived the technology to be an extremely useful asset to the Force.

**Problems and looking to the future**

Successful acquisition and integration of mapping technologies into pre-existing administrative and technical environments raised a number of issues and concerns. The survey attempted to identify and quantify a series of key problems and the degree to which they impact upon effective and efficient application of computerised crime-mapping activities. A series of topics (as identified in the US survey) were evaluated according to the extent to which they created a negative impact upon the

Constabulary and its mapping activities using a scale of 1(No problem) to 5 (Serious problem), under the following headings:

- Limited computer resources
- Limited financial resources
- Limited time
- Limited training opportunities
- Limited working knowledge of how mapping is used in the field
- Limited interest from administration
- Limited interest from support staff
- Difficulties with computer software

Primary problems and concerns for Constabularies currently engaged in mapping were identified as financial resources, time, training opportunities and limited working knowledge of how mapping is applied in the field, interest from administration and support staff creating least problems, correlating closely with output from the US survey. Non-mapping Constabularies identified computer and financial resources, training opportunities and working knowledge of how mapping is used in the field as primary concerns, interest from administration and support staff constituting minimal problems, again concurring closely with the US survey results.

### **Conclusions**

Current indication is that Governmental interest and support for is set to continue, the importance of geographic data for analysis and interpretation of crime at the local level defined in the recent Home Office report (Home Office, 2000). This Survey corroborates this sentiment suggesting that all included UK Police Forces will be engaged with computerised crime mapping within a 2-year period.

In summary, this Survey has shown both a growing influence and positive sentiment toward use and support of computer based crime mapping within UK police Forces. The overall perception of crime mapping, despite various problems and concerns is as a valuable policing tool, with both financial support from leadership and where further investment into resources is set to continue.

**PAPER 2**

**Sivarajasingam, V., J. Corcoran, D. Jones, J.A.Ware and J. Shepherd.**

“Relations between violence, calendar events and ambient conditions.” Accepted for publication in the Journal of Injury Prevention.

### Abstract

*National assault injury surveillance has identified major seasonal variation, but it is not clear whether assault injury is a seasonal problem in large cities. Relationships between community violence, calendar events and ambient conditions were investigated with reference to prospective, Accident and Emergency (A&E) derived information obtained from people injured in assaults in a European capital city (Cardiff) between 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000. Records of daily local ambient conditions included data relating to temperature, rainfall and sunshine hours. Dates of major local sporting events and annual holidays were studied. Pearson correlation coefficients were utilised to evaluate associations between variables. Overall, 19,264 assault-related A&E attendances were identified over the five-year period. Almost three-quarters were of males. Violence was clustered predominantly on Saturdays and Sundays, New Year and rugby international days. Temperature, rainfall and sunlight hours did not correlate significantly with violence ( $p > 0.05$ ). The findings indicate that injury reduction effort should be intensified at the known risk times for violence and that in a capital city/regional centre violence cannot be predicted on the basis of ambient conditions.*

**Key words:** violence, injury, association, temperature, rainfall, sunshine, calendar events

## 1. Introduction

Public health approaches to violence and injury prevention focus on risk factors [1]. They embrace the possibility that relatively unimportant risk factors and large structural factors may interact and that they provide opportunities for intervention [2]. Many studies have correlated violence with socio-demographic offender and non-offender variables, such as age, gender, race, geographic location and economic status[3]. These relatively stable criminological variables cannot explain short-term variations in violence and injury rates. In this study the focus of interest was time of violence-related injury: daily, weekly and monthly incidence of intentional injury and relationships with local ambient conditions, sporting events and annual celebrations.

## 2. Methods

### 2.1. Violence-related attendance, ambient conditions and calendar events

Prospective data relating to age, gender and attendance date for all those injured in violence who attended the only A&E department in Cardiff over a five-year period from 1<sup>st</sup> May 1995 to 31<sup>st</sup> April 2000 were studied. At all times patient confidentiality was maintained. Violence-related A&E attendance were categorised by gender and five age groups: 0-10, 11-17, 18-30, 31-50 and 50+ years. Type of software package used, flow of patients through A&E departments and stages of data capture were also studied.

Daily meteorological data on temperature (degrees centigrade), rainfall (millimetres) and hours of sunlight in Cardiff were retrieved from the local meteorological centre. Local calendar events including New Year and dates for Cardiff City football matches held at home and rugby internationals, school and bank holiday periods over the same five-year period were also identified.

## *2.2. Data analyses*

A database was assembled which formed a daily (midnight to midnight), weekly (Monday to Sunday) and monthly (first to last date) log of meteorological parameters and calendar events, and numbers of assaults by age group and gender. Simple descriptive statistics, for example cross tabulations, were used to identify both data distributions and trends. Pearson correlations were used to confirm patterns suggested by descriptive statistics. The mean number of assaults for each day of the week was combined to form a series of 60 monthly estimates. These estimates were then subtracted from the original data to create a time series of deviations from the monthly average. The time series was then fitted with a linear trend line and taking the residuals from this process, correlations between the meteorological variables and assault attendance were calculated. All data analysis was conducted using the SPSS software. A p-value of  $<0.05$  was considered to be significant.

## **3. Results**

### *3.1. Data recording in A&E department*

The Patient Administrative System (PAS) was used to enter patient data. First point of contact was with triage personnel, before registration by the receptionist at which the reason for attendance, in this case, violence-related injury, was entered (Figure 1). The PAS system included categorisation of injury as assault or accident. For every new incident a new record was created.



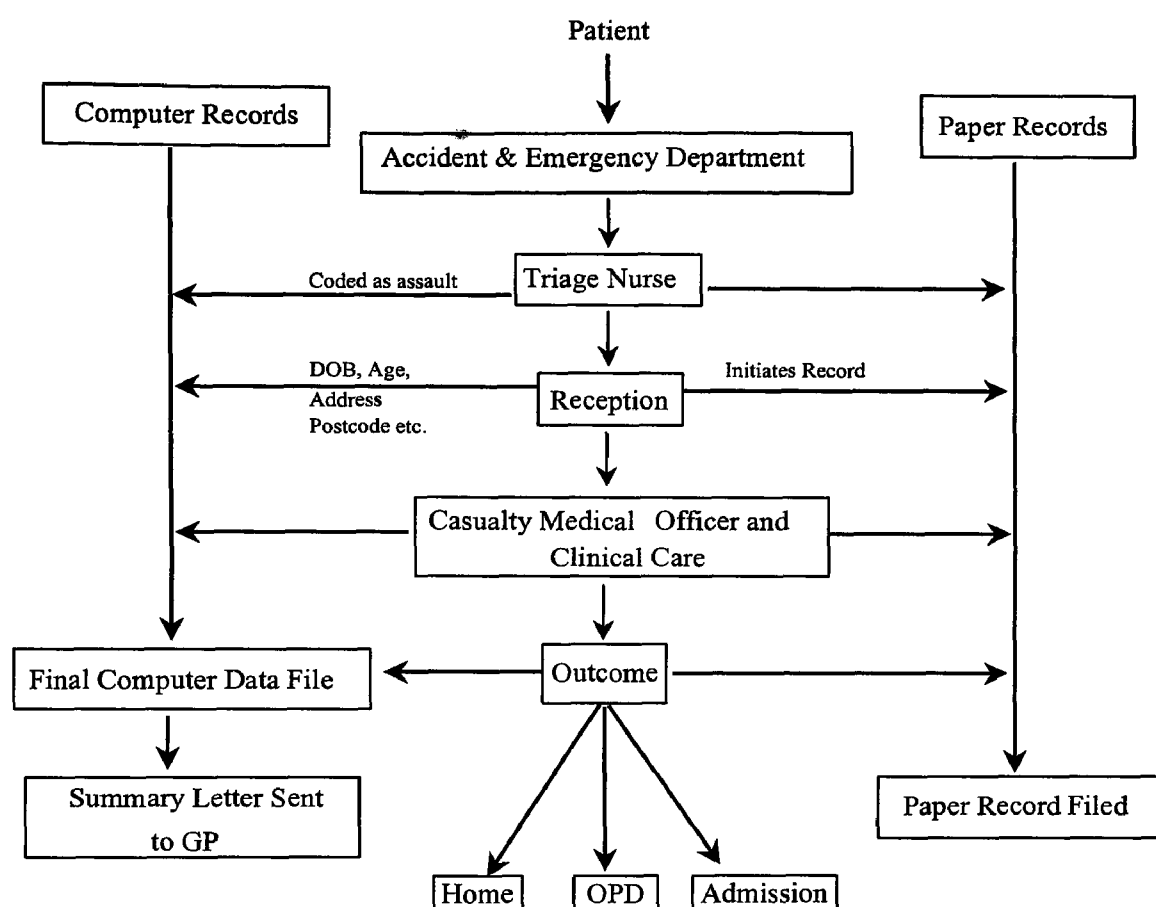


Figure 1 Flow of patients through A&E department.

### 3.2. Violence-related attendance, ambient conditions and calendar events

Overall 19,264 assaults were identified. Table 1 shows the distribution of patients by age and gender. Almost three-quarters (72%) were males. Almost half (47%) were aged 18-30 years with just over a quarter (28%) aged between 31-50 years. The mean daily attendance rate was 11 (74 per week).

Age group (years)	Male (%)	Female (%)	Total
0 to 10	471 (69.4)	208 (30.6)	679
11 to 17	2260 (71.3)	909 (28.7)	3169
18 to 30	6894 (75.8)	2204 (24.2)	9098
31 to 50	3725 (69.1)	1663 (30.9)	5388
51 +	604 (64.9)	326 (35.1)	930
Total	13954 (72.4)	5310 (27.6)	19264

Table 1: Cross tabulation of violence-related attendance by age and gender for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000

Day of the week was found to be a distinctive factor in the determination of violence-related injury attendance (Table 2 and 3). Those days where five or less violence-related injury was recorded showed a bias towards midweek days (Tuesdays, Wednesdays and Thursdays).

	Violence-related attendance						Total
	0	1	2	3	4	5	
Monday			1	4	10	20	35
Tuesday			7	10	15	27	59
Wednesday		1	3	8	14	26	52
Thursday	1	2	4	3	10	23	43
Friday			3	3	5	14	25
Saturday					4	6	10
Sunday			1	3	4	6	14
Total	1	3	19	31	62	122	238

Table 2: Violence-related A&E attendance (less than or equal to 5) by weekday for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000

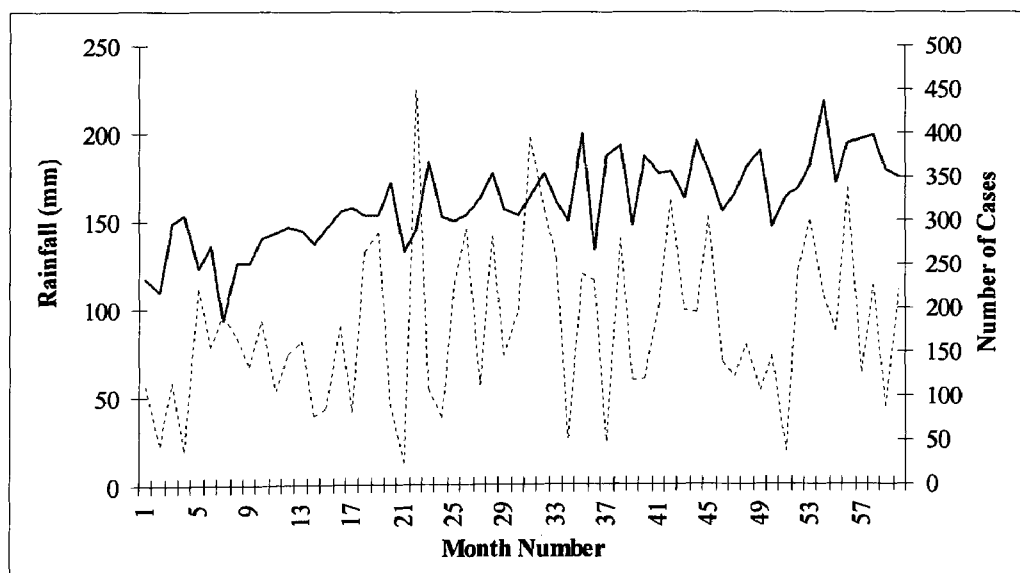
	Violence-related attendance						Total
	20 - 24	25 - 29	30 - 39	40 - 49	50 - 59	60+	
Monday	4						4
Tuesday	1						1
Wednesday	2	1					3
Thursday	3		1				4
Friday	4			1			5
Saturday	25	7	1	1		1	35
Sunday	31	6	2				39
Total	70	14	4	2	-	1	91

Table 3: Violence-related A&E attendance (greater than or equal to 20) by weekday for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000

Of the 91 days on which 20 or more violence-related injury were recorded, 74 (81%) were either on a Saturday or Sunday. Of the remaining 17 days three were New Year days, six more were in late December, one was August Bank Holiday Monday and two were rugby international days (World Cup). There were no identifiable association between the remaining five days and any events. The bias towards weekends was evident for both genders and particularly marked for those aged 18-30 years. Comparison of three monthly aggregates for spring (March, April and May), summer (June, July and August), autumn (September, October and November) and winter (December, January and

February) showed no clear seasonal pattern in violence-related attendance. Variation in attendance across age groups and gender also showed no seasonality.

Comparisons of injury data with ambient conditions revealed neither a clear seasonal pattern nor any significant meteorological influences explaining the incidence of assault (Figure 2, 3 and 4). Over the five-year period there was a steady increase (on average 2 per month) in violence-related attendance, coupled with an increase in variability. Comparison of the first and last 52-week periods showed an overall increase in attendance of males and females (40% and 37% respectively), this increase being particularly noticeable at weekends (66% for males and 68% for females). The increase was most pronounced for those aged 18-30 years. There was also a marked increase in the number of those aged 31-50 assaulted during weekends (rises of 87% for males and 70% for females). The mean daily temperature per month was 10 °C (min 3 °C, max 18 °C), 89 mm of rainfall (min 12 mm, max 225 mm) and a mean of 4 hours of sunshine (min 0.81, max 9) per day averaged over a month. Of the three parameters rainfall deviated the most from the average. Here some particularly wet months were typically preceded by especially dry periods, for example February 1997 in which 225 mm of rain was recorded followed by only 12 mm rainfall in the subsequent month. In comparison to assault-related attendance at A&E department the ambient conditions remained relatively true to annual cycles. Highest mean temperatures and sunshine hours were during the latter part of spring and summer months with lowest temperatures and sunshine hours during winter and autumn months. This annual cycle was repeated over the five-year period of the study.



**Figure 2** Changes in mean rainfall (----, millimetres) and assault-related A&E attendance (\_\_\_\_) by month for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000. Pearson correlation coefficient = 0.156 and  $p = 0.23$ .

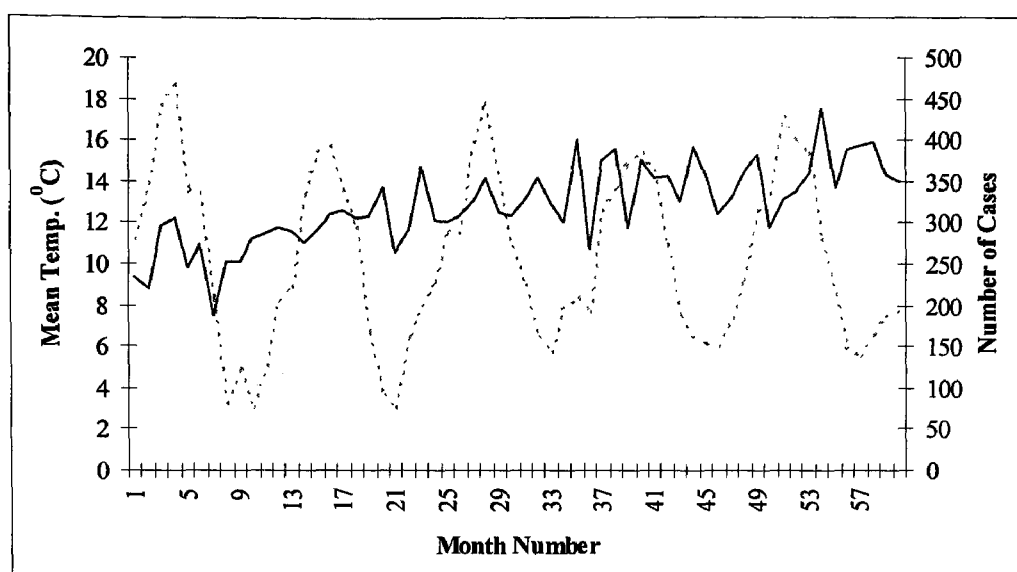


Figure 3 Changes in mean temperature (----, degrees Celsius) and assault-related A&E attendance (\_\_\_\_) by month for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000. Pearson correlation coefficient = 0.07 and  $p = 0.59$ .

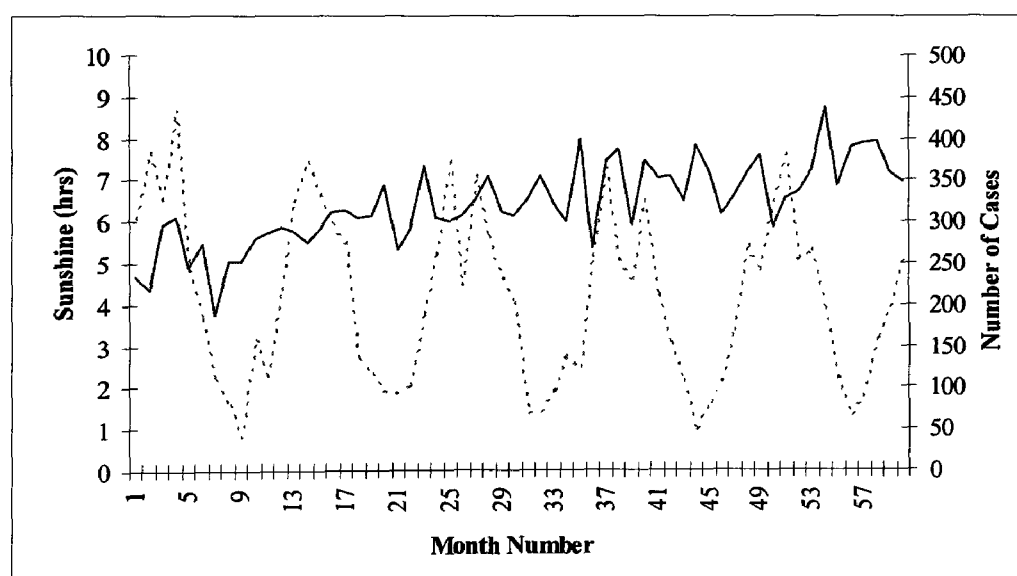


Figure 4 Changes in mean sunshine hours and assault-related A&E attendance by month for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000. Pearson correlation coefficient = -0.036 and  $p = 0.78$ .

The months with the highest assault-related attendance consistently were December and August. Analyses of calendar events indicated a positive association of both rugby internationals and the New Year celebration periods (Table 4). Out of the ten busiest days for A&E assault-related attendance over the five-year period four out of the five New Year periods were included: each year was progressively busier (114% increase from 1997 to 2000) from a violence perspective. Analyses of

assault-related injury attendance with Cardiff City football matches at home (mostly fixtures on Tuesday and Saturday) showed no correlation. Other holiday periods (school and bank holidays) also showed no association with intentional injury attendance.

Violence attendances	Day	Date	CALENDAR EVENT
62	Saturday	1 January 2000	New Year
46	Friday	1 January 1999	New Year
42	Saturday	5 February 2000	Wales v France (rugby)
34	Sunday	6 February 2000	Wales v France (rugby)
27	Saturday	23 October 1999	Wales v Australia (rugby)
37	Sunday	24 October 1999	Wales v Australia (rugby)
33	Thursday	1 January 1998	New Year
30	Saturday	18 December 1999	Unknown
29	Wednesday	1 January 1997	New Year
29	Saturday	11 March 2000	Unknown
28	Saturday	15 March 1997	Wales v England (rugby)
27	Saturday	21 March 1998	Unknown

Table 4 Top twelve days for violence-related A&E attendance by calendar event for period 1<sup>st</sup> May 1995 and 30<sup>th</sup> April 2000

#### 4. Discussion

Few theories have considered the immediate situational factors that may help explain why a particular criminal event occurred at a particular time. Situational approach to crime as well as rational choice theory, suggest that immediate crises, events and conditions are important factors in the offenders decision to commit a crime [4]. The routine activity theory attempts to examine the relationship between climatic variables and criminal behaviour [5]. This theory suggests that individual activity and daily habits are rhythmic, and consist of patterns that are repeated over time. However, changes in the surrounding environment may result in changes in behaviour and activities. For example, during pleasant weather, people tend to spend more time out doors, resulting in greater opportunities for personal interaction and increased availability of victims.

In this study we have confirmed that, from the health service perspective, Saturdays and Sundays were consistently busiest in terms of the number of individuals seeking treatment following violence-related injuries over a five-year period. This was evident for both males and females for all the age groups studied, particularly for those aged 18-30 years. It has been suggested that spare-time activity is a significant factor in determining the risk of assault injury [6]. Those aged 18-30 years are likely

to spend proportionately more of their spare-time, for example during weekends, in licensed premises and outdoors in the street where there are higher risks of violence and injuries. Daily fluctuations in the number of those injured in violence over such a long period have not been previously reported. The present study also showed no significant seasonality in assault-related A&E attendance for both males and females over the five-year period. The attendances during the winter months (December, January and February) were not different to attendance during spring (March, April and May), summer (June, July and August) and autumn (September, October and November) months. A previous study evaluated assault-related attendance in 58 A&E departments in England and Wales over the same five-year period and reported significant violence seasonality (assaults least frequent during winter and autumn months, and most frequent during spring and summer months) [7]. It was postulated that there may be links between relatively high temperatures, hours of daylight and high injury rates, and between relatively low temperatures, hours of daylight and low injury rates. However, in the study reported here the ambient conditions remained relatively true to annual cycles with rainfall showing the greatest variability.

The result of this study has shown that there was no correlation between violence-related injury A&E attendance and ambient conditions. Previous studies however, have identified positive linear relationship between temperature and assault. One such study showed that temperature variables were the most important daily predictors of assaults in comparison to air pollution and barometer pressure [8]. They reanalysed archival data of serious civil disorders in America between 1967 and 1971 and found that the relationship between temperature and crime became linear, so that the probability of a riot increased steadily with ambient temperature up into the mid-90s. Another study found that daily mean temperature was a significant predictor of the rate of domestic complaints over a two-year period [9]. A positive linear relationship between the monthly mean temperature and the monthly number of crisis calls received by battered women's shelters in five locations in the United States over a two to three year period was also reported [10]. In this respect, the results of the study reported here are not consistent with previous research correlating temperature and assault. The relationship between sunlight hours and assault is far less clear. A study conducted in Dayton, Ohio, found hours of daylight and assault to have a significant, positive linear relationship [9]. However analyses of annual change in photoperiod by the latitude of each location showed no relationship between photoperiod and homicide, assault, rape or robbery in any of the locations studied [10]. Changes in temperature and sunlight hours tend to covary, hence it is difficult to separate their effects. Similarly, the relationship between rainfall and assaults is uncertain because the results are inconsistent. One study reported a negative correlation between precipitation and assaults [11] while another [12] reported a significant positive correlation. However, it is important to point out that all the above studies utilised *reported* (police) assault data and differ from the A&E department *recorded* assault

injury data evaluated in this study. It is possible that temperature, sunshine hours and rainfall affect police reporting of assaults in some way but has no effect on assault-related A&E attendance.

Calendar events have also been used to explain seasonal variations of crime [13]. Calendar events describe particular points and periods in the year, public or school holidays, pay dates and sports events. Each has the potential to describe variations in criminal activity where a fluctuation in a particular type of crime is a function of a temporal occurrence. Association of assault-related injury and New Year and dates of rugby internationals found in this study are interesting. These calendar events are likely to attract large numbers of people at selected venues. It is well known that violence in city-centres is often concentrated in only a few streets, which usually contain a relatively large number of public houses such as pubs, clubs and discotheques [14]. The mix of large numbers of people before, during and after these events both within and outside public houses and invariably the role of alcohol could explain the greater number of injuries seen at A&E departments on these dates.

The lack of seasonality in assault-related A&E attendance at Cardiff found in this study and seasonality in assault-related attendance reported in an earlier national study may reflect differences in frequency of calendar events in cities and towns across England and Wales. It is possible that in those cities and towns where violence and injuries are seasonal the particular calendar events correlating with high assault-related A&E attendances are also seasonal. In Cardiff being a capital city of Wales it is possible that violence is evenly spread because of the comparative frequency of events spread throughout the year. This might not be similar to non-capital cities across England and Wales.

From the public health perspective the finding that violence-related injury is correlated to calendar events, especially New Year periods and Rugby Internationals and is not correlated to ambient conditions (temperature, rainfall and hours of sunlight) are important. The previous finding of a significant correlation between national violence-related A&E attendance and seasonality may simply reflect calendar events and not the effects of ambient conditions as postulated [7]. The findings indicate that injury reduction effort should be intensified at the known risk times for violence and that violence cannot be predicted on the basis of ambient conditions.

### **Acknowledgements**

V. Sivarajasingam was the principal author of the paper and retrieved Accident & Emergency department data. J. Corcoran and D. Jones retrieved data from the Cardiff meteorological centre, on holidays and sport fixtures and carried out most of the analysis. A. Ware and J. Shepherd designed and supervised the study. V. Sivarajasingam will act as overall guarantor.

---

**References**

- [1] Mercy JA, O'Carroll PW. New directions in violence prediction. *Violence vict* 1988; 3: 285-301.
- [2] Shepherd JP, Farrington DP. Assault as a public health problem. *J Roy Soc Med* 1993; 6: 89-92.
- [3] McLoyd VC. The impact of economic hardship on black families and children: Psychological distress, parenting and socio-economic development. *Child Dev* 1992;10: 9-15.
- [4] Clarke RV, Cornish DB. 'Modelling offenders' decisions: A framework for research and policy. In Tonry M, Morris N. *Crime and Justice*. 4<sup>th</sup> ed. Chicago: University of Chicago Press 1985 147-85.
- [5] Cohen LE, Felson M. Social change and crime rate trends: A routine activity approach. *Am Socio Rev* 1979; 44: 588-608.
- [6] Smith SJ. Victimisation in the inner-city. *Br J Criminol* 1982; 22: 386-402.
- [7] Sivarajasingam V, Shepherd JP, Matthews K, Jones S. Trends in violence in England and Wales 1995–2000: an accident and emergency perspective. *J Pub Health Med* 2002; 24(3): 219-226.
- [8] Carlsmith JM, Anderson CA. Ambient temperature and the occurrence of collective violence: A new analysis. *J Personality and Soc Psychol* 1979; 37: 337-44.
- [9] Rotton J, Frey J. Air pollution, weather and violent crimes: Concomitant time-series analysis of archival data. *J Personality and Soc Psychol* 1985; 49: 1207-1220.
- [10] Michael RP, Zumpe D. An annual rhythm in the battering of women. *Am J Psych* 1986; 143: 637-40.
- [11] Feldman HS, Jarmon RG. Factors influencing criminal behaviour in Newark: A local study in forensic Psychiatry. *J Foren Sci* 1979 ;24: 234-9.
- [12] Perry JD, Simpson ME. Violent crimes in a city: Environmental determinants. *Environment and Behaviour* 1987; 19: 77-90.
- [13] Harries KD, Stadler SJ, Zdorkowski RT. Seasonality and Assault: Explorations in Inter-Neighbourhood Variation, Dallas 1980. *Annals of the Association of American Geographers* 1984; 74(4): 590-604.
- [14] Shepherd JP, Brickley M. The relationship between alcohol intoxication stressors and injury in urban violence. *Br J Criminol* 1996; 36: 546-66.



**PAPER 3**

**Corcoran, J. and B. Bowen Thomson**

“New Insights into Community Safety: Application of the HASCADE Model.” Accepted for publication in the British Journal of Community Justice.

A summary of this paper was presented at the Managing the Strategy Conference, Cardiff 2003.

**Abstract**

*Through evaluation of current national approaches to community safety a non-prescriptive model is presented through which sustainable strategic decisions can be founded. The HASCADE model (Holistic Approach to Strategic Crime And Disorder Evaluation) focuses on mainstreaming community safety at the local level through providing an explanation of community dynamics. Central to HASCADE is the establishment and consolidations of inter and intra- agency collaborations coupled with mapping and statistical techniques. Relating to practical experience, this paper articulates the issues, methodology and future application of HASCADE model implementation.*

**Keywords:** Community safety, mainstreaming, Geographical Information System (GIS), holistic.

### Introduction

The control of crime and disorder is not solely the remit of the police nor are relevant data solely to be found in their records. If one is to understand fully the dynamics and needs of a region or community there is a need to consult additional data, sourced from a range of public and commercial organisations, agencies and authorities at the local level (Graham *et al.* 1998). This premise has partly driven the promotion and development of local partnerships to guide and facilitate data sharing and the data collation, aggregation and analysis process.

The need to minimise community disorganisation through tackling crime and disorder issues was formalised through the Crime and Disorder Act 1998. The Act places a legal obligation on the local authority and police to work in tandem to develop, publish and implement 3-year strategies to tackle crime and disorder based upon findings of a local crime and disorder audit. In addition the Act stipulates the necessity to work with other key agencies, including health, education, business and voluntary sectors (see Sections 5-7, Crime and Disorder Act, 1998).

The development of strategies is assisted by an audit of local crime and disorder problems conducted every 3 years. At the time of writing the local partnerships, known as Crime and Disorder Reduction Partnerships (CDRPs), have recently completed their second audit to direct local policy and strategy until 2005. The aim of the audit to provide a current snapshot of local crime and disorder issues. Its main thrust is to offer insight into the scale and scope of crime and disorder over time (commencing with appraisal of achievements since the previous audit) through drawing together multiple agency data to inform future guidelines and strategy. It does not constitute a traditional audit in the sense of attempting to produce costings of crime, more broadly it aims to '*manage performance effectively, pooling information to identify and analyse problems*' (Audit Commission 2002: 41).

The important elements of strategy development rely heavily upon the outcome of the analysis within the audit. '*A model strategy is one that: Is analysis driven. Explicitly understands the process by which the plan is to reduce crime. Employs interventions that are mutually advantageous. Understands the importance of sequencing*' (Curtin *et al.* 2001: 22).

It appears then, that the analysis plays a crucial role in strategic development. To develop a holistic strategy, the analysis of multiple datasets is important, particularly if that analysis indicates potential for mutually advantageous interventions. If the data being analysed for the Crime and Disorder Audits primarily focus upon crime and disorder incidents, then it is difficult to establish a strategy that will incorporate the multitude of partner agency objectives. This may be of particular importance when considering the potential role of audit and strategy in informing and facilitating compliance with Section 17 of the Crime and Disorder Act 1998. Section 17 of the Act states '*without prejudice to any other obligation imposed on it, it shall be the duty of each authority...to exercise its various functions with due regard to the likely effect of the exercise of those functions on, and the need to do all that it reasonably can to prevent, crime and disorder in its area*' (Crime and Disorder Act 1998: Section 17-1).

The remainder of the paper discusses current perspectives and their limitations for conducting an audit and importance of a geographically orientated approach. The paper concludes by articulating a new approach to auditing crime and disorder through the adoption of an holistic perspective.

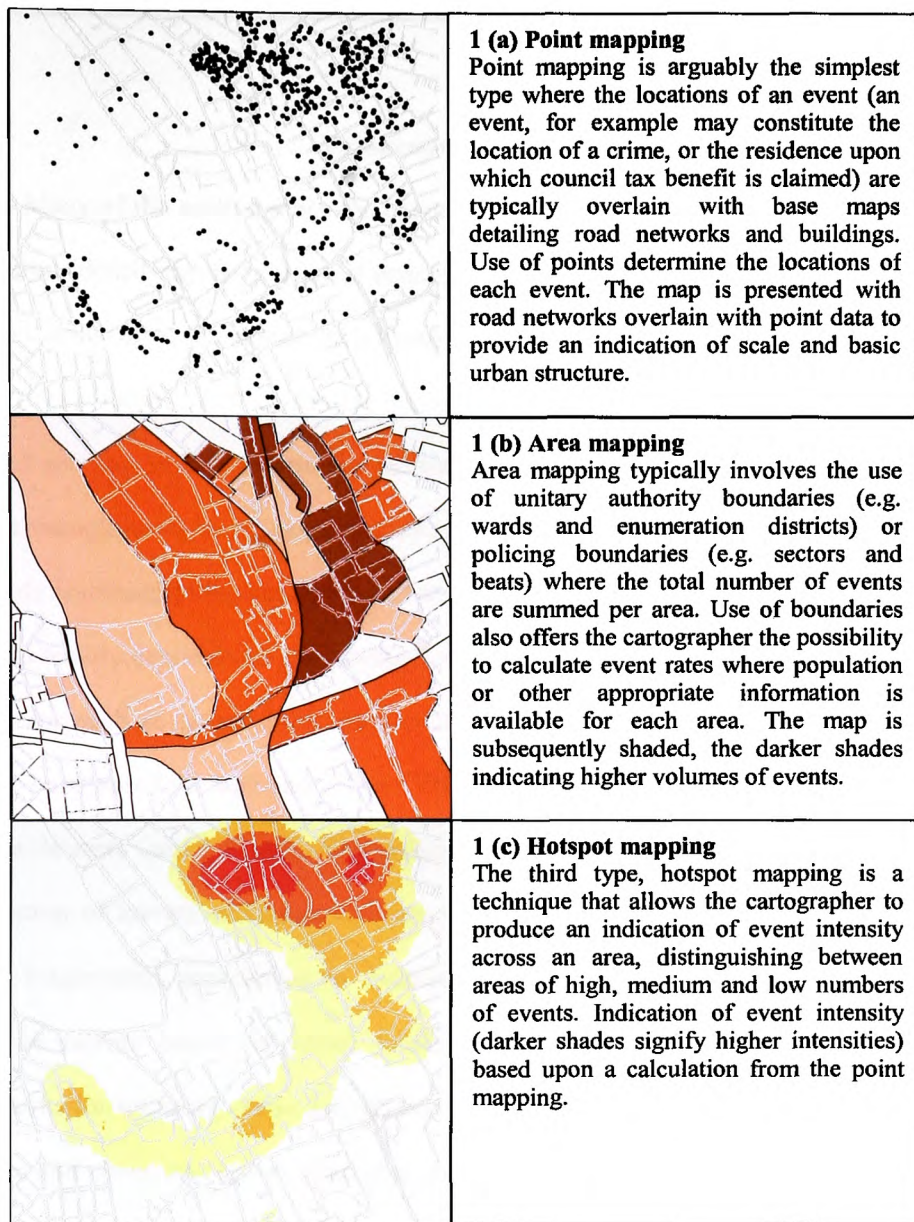
### Geographical Analysis

The importance of locale in the understanding of criminal dynamics has been the focus of a large and growing body of research (Anselin *et al.* 2000) that arguably has been both supported and fuelled by recent technological advances in the field of computerised mapping and Geographical Information Systems (GISs). GISs offer the potential to automate many processes that were once confined to manual procedures, making the production of mapping both a less time consuming and less resource intensive process. For example, production of a map depicting locations of crimes in a region prior to the GIS era would have demanded a skilled cartographer devoting many hours to manually scribing the information on to paper map sheets. Today GISs offer the potential to automate this process, whereby the cartographer is able to reproduce many alternative outputs in a fraction of the time. This has

opened up a whole gamut of applications that GIS can be applied to, one of which is the mapping of crime and disorder.

The past 5 years have seen a marked growth in the use of computerised mapping systems by police forces in both the UK (Corcoran and Ware 2002) and US (Mamalian and LaVigne 1999), a trend set to continue. Amongst the numerous tools that GIS offer to the cartographer, three main types of outputs are commonly used to depict locations or volumes of criminal activity: point, area and hotspot mapping (Figure1)

The ability to pinpoint and visualise the precise locations of events has seen the promotion of micro-level analysis that has become of particular interest to situational crime prevention (SCP) programs and has seen promotion by the Government (Home Office 2000). Micro level analysis can prove successful in such programs where the objective is to uncover the specifics of a locale in explanation of its propensity toward observed events e.g. a series of houses within a neighbourhood particularly subject to burglary. However, in the context of the audit use of such techniques places a large demand upon each partner to provide full address information from which the data can be geocoded (translated into an x and y co-ordinate) to the fine scale typically demanded by SCP programs.



**Figure 1 Common types of mapping (a) point (b) area and (c) hotspot mapping**

In addition, the role of the audit is usually to provide an overview of a whole local authority area, thus fine resolution analysis is arguably not the primary objective. Moreover, the audit should put in place a series of analyses that is capable of identifying the broad issues and then is able to direct the identified partners who should consider a targeted response. At this stage a micro level analysis could take place to isolate the specific issues (e.g. vulnerable houses and common modus operandi) to ensure a correct application of preventative measures (e.g. a lock fitting scheme and advice).

### Perspectives on Guidance

Many of the multi-agency data sources that appear to have been used by Partnerships focus upon identifying geographical concentrations of crime and disorder. Alongside police data numerous other databases have been recommended as useful for audits (Hough and Tilley 1998). Often such databases focus upon the type of crime and disorder that has occurred and can further inform the nature and level of crime and disorder within an area. Caution though needs to be exercised when considering data that only refers to a section of the whole community: housing data on vandalism or anti-social behaviour, for example, may not be so readily available on the private-rented or home-owner sector. The value of analysing information such as geographical areas exhibiting high densities of offenders and census data has also been noted (Hough and Tilley 1998). Yet, less emphasis has been given to analysing *multiple* datasets aimed at indicating potential causes of crime and disorder, whether this be the location of known offenders or other risk and vulnerability factors. Such analysis may inform longer-term approaches to reducing risk and vulnerability factors and increasing protective factors based on social forms of crime and criminality prevention. In their frameworks for reducing crime, Hough and Tilley note that *'identifying the causes of crime is complex and there are many competing theories. Analysing particular crime problems for preventive purposes is, however, more straightforward...'* (Hough and Tilley 1998: 36). The frameworks they advocated are based on SCP including Rational Choice Theory (Clarke and Felson 1993) and Routine Activities Theory (Cohen and Felson 1979).

SCP consists of three measures aimed at reducing offending opportunities. These measures *'...(1) are directed at highly specific forms of crime, (2) involve the management, design or manipulation of the immediate environment in as systematic and permanent way as possible, (3) make crime more difficult and risky, or less rewarding and excusable as judged by a wide range of offenders'* (Clarke 1997: 4). For SCP to be incorporated into delivery, detailed micro-level analysis that distinguishes between specific crime categories is required. This involves detailed and accurate information that is not always readily available or

accessible from agencies. Of the multi-agency datasets that record crime and disorder incidents, police data are probably the most (although not 100%) accurate and easily accessible. Given the relative ease of accessibility, it would still be time consuming and costly to extract the relevant detail required. More recent guidance on crime audits stresses the importance of not relying solely upon police data, noting issues regarding under-reporting and the risk of simply mirroring existing policing priorities (Fox and McManus 2001). Yet, an approach to reducing crime and disorder based primarily upon SCP typically relies upon such micro-level analysis.

Prior to conducting such resource intensive micro-level analysis based within a geographical area, it is important to ascertain where such analysis should be targeted. Considering the spatial composition of multi-agency datasets, chosen to reflect related criminological and vulnerability issues, would assist evidence based decision-making and targeting different community safety approaches to the most appropriate areas. Recognition has been given to considering information and databases such as socio-demographic information and school exclusion data, but little attention has been given to how this information could be analysed alongside crime and disorder incidents to inform strategy development (Hough and Tilley 1998; Fox and McManus 2001).

Much evidence exists that details the benefits of the application of various SCP techniques to crime and disorder problems (Clarke 1997; Felson and Clarke 1998). A crime and disorder reduction approach based solely upon the application of SCP techniques is likely to elicit a variety of outcomes, some of which will be successful, some unsuccessful, resulting in unbalanced and possibly short-term outcomes that are at risk of excluding the most vulnerable community members. There is a risk that such an approach may result in social exclusion, individuals being '*excluded from citizenship*' (Young 2002: 465). Similarly, as noted by Shapland (2000), criticisms of SCP approaches note that they are at risk of: increasing exclusion, reinforcing a fortress society, threatening civil liberties, inconveniencing law-abiding citizens and resulting in victim blaming. It seems therefore important that crime and disorder audits are able to identify multi-agency issues that can



inform strategy and policy direction on crime and disorder issues, alongside the community dynamics within an area. Simply focussing upon micro-level analysis, by locating singular crime and disorder events, to aid SCP approaches, obstructs attempts to identify the existence of potential risk and protective factors within an area. Thus a crime and disorder audit based primarily upon the location of crime and disorder events risks advocating an unbalanced, socially exclusive approach that does not consider the wider criminological implications.

To facilitate the mainstreaming of community safety, a more inclusive and holistic approach to reducing crime and disorder and promoting community safety would be advantageous, as a wider variety of agencies would be able to relate to such a discourse. Clarke (2000) notes that most criminologists are not enthused with SCP, therefore it is understandable that some partners, who play an important function in promoting community safety do not prioritise crime and disorder reduction as they may have difficulty relating their responsibilities with SCP ideology. Thus, it appears that a variety of approaches would facilitate multi-agency collaboration and a true partnership approach to the reduction of crime and disorder and the promotion of community safety. As noted by Smith (2000: 172), *'it is not necessary to argue that SCP is generally superior to the wide variety of other methods of crime prevention, especially since many of these other methods have different, often wider, objectives, and are aimed at different target groups.'*

### **The HASCADE Approach**

Modelling techniques to direct, monitor and evaluate community initiatives demands a holistic approach to be adopted whereby a range of local information is considered and analysed in an appropriate manner. HASCADE provides a flexible framework for analysing multiple datasets, including data from a variety of sources. Prior to data being requested, consideration is given to what those data represent. The HASCADE model aims to identify community vulnerabilities as well as crime and disorder events. These community vulnerabilities can provide an insight into the community composition across a region.

Crime and disorder as recorded by the police arguably constitutes only a partial descriptor of the community (Shepherd *et al.* 1989), as issues relating to crime and disorder affect all sectors of the community. Even so, an unequal distribution of crime risk exists – *'40% of crime takes place in 10% of neighbourhoods. Many now recognise the need to target the neighbourhoods experiencing the worst crime problems with a comprehensive package of crime reduction measure'* (Browne *et al.* 2001: 2). HASCADE identifies concentrations of community vulnerabilities, crimes and disorder, thus assessing whether community vulnerabilities exist within geographical areas that exhibit the highest levels of crime and disorder. The following table (Table 1) highlights the datasets used and how they may relate to community vulnerabilities:

Vulnerability	Related Datasets
Economic Poverty	Council Tax Benefit Claimants
Peer Criminality	Police Reprimand, Final Warning, Sentence - Youth Offending Team (YOT) plus supervision and unsupervised data – Probation
Family Criminality	YOT and Probation data, as above
Lack of Commitment to school	School Exclusions
Future Risk of Lack of Educational Attainment	School Exclusions
Social Exclusion	All of the above, Looked After Children and Benefit Fraud
Crime and Disorder	Police incident figures, Community Safety Questionnaire Benefit Fraud

Table 1 Description of datasets and their relationship to vulnerability.

Table 1 is not exhaustive, but places importance upon the representation of data, relating various datasets with possible explanation. This approach is in contrast to a data warehousing approach that contains numerous datasets collated simply because they exist and are collectively considered. Such an approach may increase the risk of *'rubbish data in:*

*rubbish analysis out'* (Radburn 2000: 1.05). A targeted approach to data selection enables an increased understanding of the difficulties and inconsistencies that are inherent in all datasets.

To access the data both a top-down and bottom-up approach were used. Using a top-down approach, a data list was developed that reflected crime, disorder and vulnerability criteria, permissions and agreement obtained from members of the CDRP. Key individual partners who worked directly with each data set were then identified and approached. This enabled the incorporation of a bottom-up approach into the design. These key individual partners heightened awareness of inconsistencies within the data that may not be publicly or commonly known. Specific data issues could then be identified. This knowledge was used to select and discard the final datasets. As not all data were available in the required format some datasets; such as housing voids, to represent disorganised communities; had to be excluded from the model. Timescales also produced difficulties in obtaining some of the datasets, (such as noise nuisance) and concerns regarding the reliability of certain recording practices also resulted in some datasets being ruled out (for example, unauthorised and authorised absences from schools).

### **Analysis for HASCADE**

Analysis of the data was divided into two key stages, GIS and statistical, resulting in the production of a summary map. The techniques used in each stage are discussed in turn.

### **GIS**

The initial task was one of visualising all data within the GIS. Using the spatial reference (i.e. the postcode) provided by all Partners all data was geocoded, a process to assign an x and y co-ordinate to each record. Using Code-Point (a Gridlink® Product) the x y co-ordinate referenced a postcode unit (an area on average containing 15 addresses) thus helping to allay concerns regarding privacy, but still offering an acceptable level of spatial accuracy to permit analysis.

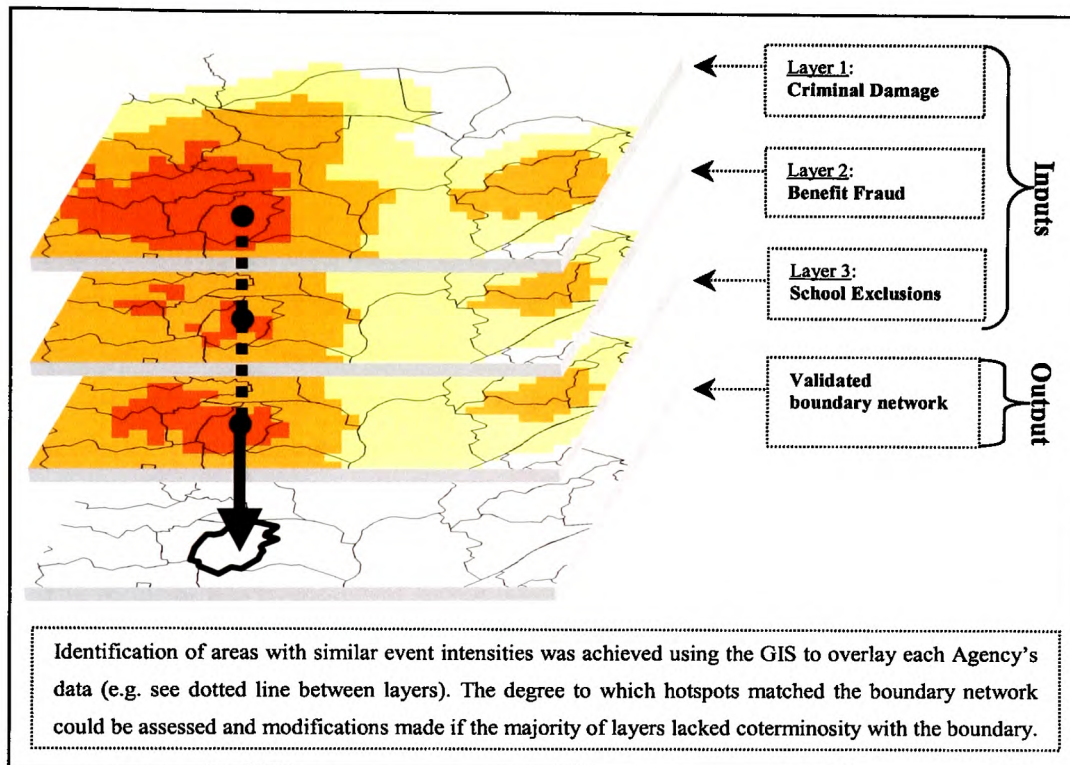
Once geocoded, data sets were visualised in the GIS at point level overlain with various street and boundary information (e.g. roads networks and building foot prints) to provide context. In many cases the volume of mapped events created visualisation problems at point level where locations with numerous events mapped appeared as a single occurrence (each event mapped over the next). To improve visualisation hotspot maps were created enhancing the depiction of event intensity across the region. This was of particular relevance for the visualisation of high volume events, particularly within city centre districts, where numerous events were typically located at the same geographic locality.

To provide greater context to the underlying population geography to which events related it was necessary to generate a series of aggregate maps. Using a modified boundary network based upon the 1991 Census enumeration districts (containing 1999 population estimates), each Partner's data were aggregated to the new framework. The number of events contained within each region was calculated in addition to their respective rates based upon each area's population.

Thus far the GIS analysis focused on building a visual understanding of event distribution across all Partners' data, following which it was possible to commence a primary identification of Priority Geographical Areas (PGAs) on the strength of point, hotspot and aggregate mapping. Aggregate outputs offered an indication of vulnerability at broad neighbourhood scale, with point and hotspot mapping identifying more specific sub-neighbourhood localities internal to these regions.

The final aspect was to confirm and quantify these visual perceptions. This was achieved through the development of a common boundary network that best described all data. Overlaying hotspot mapping of each Agency's data (as this best describes event distribution) over the boundary network, previously described, its closeness of fit was determined. The aim here was to achieve the closest and most feasible representation between network and event distributions (Figure 2). Where there was an identified lack of coterminosity the boundary network could be modified to provide a closer fit to the Agency data. Typically a modification may include aggregating two or more regions together. It was

evident that a perfect match between all data and the boundary network could never be achieved, however it was deemed acceptable where the majority of data were well represented.



**Figure 2 Method of validating the boundary network.**

### Statistical

Extracting the event rates for each area within the validated boundary network, statistical analysis targeted the determination of correlation between the various data sets to reinforce relationships identified through spatial inspection. Using the Pearson's correlation test provided a measure of linear association, where significant relationships at the 0.01 confidence level were identified. Establishing significant statistical linkages together with identification of PGAs assisted the formulation of strategy through developing a fundamental understanding of criminogenic indicators.

Thus, following spatial inspection, the identification of a neighbourhood area as a vulnerable locale for events such as school exclusions and youth offending can be supported or questioned through the identified statistical linkages. The resultant of this hypothetical

scenario would be to offer a target set of organisations to jointly direct interventions in the specified area. In addition such output and evidence produced through the implementation of this framework reinforces the necessity of Partnerships, supporting the current drive towards joined-up government working practices.

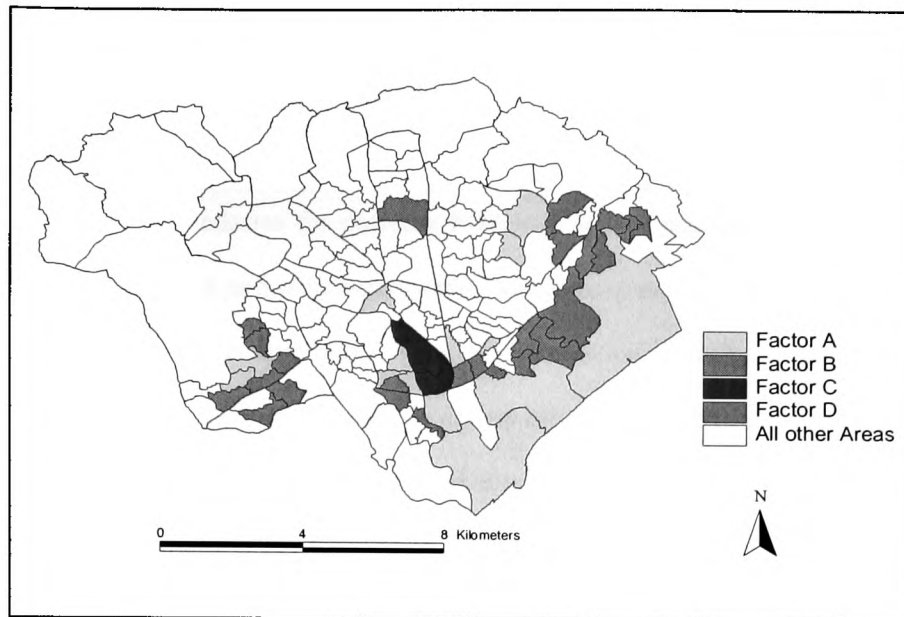
Based upon visual inspection and statistical associations the final stage involved a categorisation of the CDRP region into a series of PGAs. For the HASCADE model a heuristic was developed whereby PGAs could be identified on various levels. Four categories were used to encapsulate the variety of issues that in turn would formulate the priorities. The FACTOR A could be considered as primary priority geographical areas. FACTOR B and FACTOR C considered as secondary priority geographical areas and finally, FACTOR D regarded as tertiary priority geographical areas (Table 2). To a certain extent, a form of sliding scale of vulnerability can naturally apply, moving from Factor A through to Factor D. Yet, through HASCADE it is also possible to identify small areas, which, when considered holistically within the CDRP area may indicate inter-relationships existing between small areas and across Factors, identifying issues and areas requiring further explanation.

As noted in the factor Map, areas that exhibit high levels of vulnerability (Factor D) are not necessarily the same areas that exhibit high levels of crime and disorder. This possibly indicates the existence of protective or risk factors. Similarly, it is not inevitable that areas exhibiting high levels of crime and disorder are the same areas where high numbers of offenders reside. HASCADE further informs a holistic approach to crime and disorder reduction in its attempts to distinguish between the variety of issues that may impact upon an area. Thus, to achieve sustainable crime and disorder reduction, techniques can be applied that address issues specific to the particular area, aiming to improve cost effectiveness.

Factor	Description of Geographical Areas	Type of response
FACTOR A	Exhibiting several types of crime, disorder <u>plus</u> community vulnerability	A balanced approach: recourse to criminal justice system, situational crime prevention plus prevention of future criminality
FACTOR B	Exhibiting one specific type of crime, disorder <u>plus</u> several types of community vulnerability	A balanced approach: recourse to criminal justice system, situational crime prevention plus prevention of future criminality
FACTOR C	Exhibiting several types of crime, disorder <u>but</u> no community vulnerabilities	Recourse to criminal justice system and situational crime prevention methods
FACTOR D	Exhibiting no crime and disorder, <u>but</u> several types of community vulnerabilities	Support to: promote community safety and prevention of criminality. In-depth analysis to highlight any existing good practice

Table 2 Description of factors and type of response.

Using a count system each of the areas within the boundary network was classified according to the criteria detailed in Table 2. The resulting map offered a compact, easily interpretable summary of all Partners' data and a foundation from which future policy could be derived (Figure 3).



**Figure 3 Factor Map showing PGAs.**

### Discussion

HASCADE facilitates the construction of a balanced approach to crime and disorder reduction and the promotion of community safety. It aims to inform multi-agencies of their potential role in reducing crime and disorder and increasing community safety.

The model provides a flexible, organic framework that enables an improved method by which mainstreaming can evolve. Mapping processes coupled with statistical techniques can result in a powerful visual and explanatory tool. Together, these techniques aid effective inter and intra-agency collaboration targeted at achieving common goals through agreed approaches. Through HASCADE, a series of techniques are offered whereby a balanced approach can be implemented. Advocating such an approach reduces inter-and intra-agency conflict. A balanced approach supports the realisation of the value of community safety and its contribution to sustainable joined-up government.

HASCADE is an inclusive approach able to incorporate a larger variety of data. The model does not rely upon detailed micro-level analysis; thus, the resolution of spatial data required is less specific. GIS is a new requirement for many agencies, thus, in common with other GIS approaches, data may not be readily available in the requested format. This may be



particularly true for voluntary agencies, such as social landlords and domestic violence groups. Subsequently, this reduces the reliance upon each agency to collect pin point (full address) level information.

The use of boundaries for defining PGA's is one of the limitations of HASCADE. Some crime and disorder and community safety issues may not conform to such boundaries, for example some virtual communities. There is also a danger that the aggregated areas used within HASCADE results in an ecological fallacy (Robinson 1950), the misclassification of areas so that they fail to accurately represent the true underlying characteristics. Similar to other methods of analysing information using GIS, outliers will be lost with HASCADE. Such scenarios may require a more detailed statistical appraisal of the data.

The inclusion of HASCADE, with its ongoing data analysis, into the mainstream community safety process should continually inform crime and disorder reduction. Currently, the Crime and Disorder Act (1998) requires the completion of an audit of crime and disorder every three years. For the audit process to be ongoing the Implementation of regular HASCADE cycles is required, enabling CDRPs to provide a broad evaluation of the overall strategic aims and to ensure that initiatives are targeting the most appropriate areas.

### **Summary**

HASCADE is easy and cheap to implement, as it is not dependent upon high levels of customisation. Thus resulting a sustainable, replicable and flexible crime and disorder reduction model.

A move away from the more traditional approaches utilised to reduce crime and disorder where interventions are targeted at areas exhibiting high crime and disorder levels, based upon the underlying assumption that marries high crime and disorder areas with high vulnerability. HASCADE shows that this is not always true, by offering explanations for intervention based upon multiple datasets.

The HASCADE model can assist with promoting and supporting the mainstreaming of community safety, as both statutory and non-statutory partners are incorporated into the process.

HASCADE endorses crime and disorder reduction interventions based upon ongoing analysis and evidence-based practice. Thus, the model can be more regularly applied than the required three-year audit cycle. Similarly, the model can be extended to a national framework.

### References

- Anselin, L., J. Cohen, D. Cook, W. Gorr and G. Tita (2000). 'Spatial analyses of crime', in D. Duffee (ed.), *Criminal Justice 2000: Measurement and Analysis of Crime and Justice*. Washington, DC: National Institute of Justice: (4): 213–262.
- Audit-Commission (2002). *Community Safety Partnerships: Learning from Audit, Inspection and Research*. Audit Commission.
- Clarke, R. V. and M. Felson, Eds. (1993). *Routine activity and rational choice. Advances in Criminological Theory*. New Jersey: Transaction.
- Clarke, R. V. (1997). *Situation crime prevention: Successful case studies*. Albany, N.Y: Harrow and Heston.
- Clarke, R. V. (2000). 'Situational prevention, criminology and social values', in A. Von Hirsch, D. Garland and A. Wakefield (eds.) *Ethical and social perspectives on situational crime prevention*. Oregon: Hart Publishing: 97-112.
- Cohen, L. E. and M. Felson (1979). 'Social change and crime rate change: A routine activities approach.' *American Sociological Review* 44: 588-608.
- Corcoran, J. and J. A. Ware (2002). 'Helping with enquiries'. *GeoConnexion*. 1(3): 36-40.
- Browne, C. and R. Chadwick (2001). *Reducing Neighbourhood Crime*. Crime Concern.
- Curtin, L., Tilley, N., Owen, M., Pease, K. (2001). *Developing crime reduction plans: Some examples from the Reducing Burglary Initiative*. London, Home Office.
- Fox, C. and McManus, J. (2001) *The NACRO Guide to Crime Audits*. London: NACRO.
- Felson, M. and Clarke, R. V. (1998). *Opportunity makes the thief: Practical theory for crime prevention*. Police Research Series paper 98. London: Home Office.
- Graham, J., Ekblom, P., Pease, K. (1998). *Reducing offending: An assessment of research evidence on the ways of dealing with offending behaviour*. Research Study 187. London: Home Office.
- Home Office (1998). *Guidance on Statutory Crime and Disorder Partnerships: Crime and Disorder Act 1998*. London: Home Office Communication Directorate.

Home Office (2000). Review of Crime Statistics: A Discussion Document. Research Statistics Directorate. London: Home Office.

Hough, M. and Tilley, N. (1998). *Auditing crime and disorder - Guidance for local partnerships*. Crime Detection and Prevention Series Paper 91. London: Home Office.

Mamalian, C. A. and LaVigne, N. G. (1999). *The Use of Computerised Crime Mapping by Law Enforcement: Survey Result*. Washington: National Institute of Justice.  
<http://www.ncjrs.org/pdffiles1/fs000237.pdf>.

Phillips, C., Considine, M., Lewis, R. (2000). *A review of audits and strategies produced by crime and disorder partnerships in 1999*. Briefing Note 8/00. London: Home Office.

Radburn, S. (2000). *Data exchange & crime mapping*. East Midlands: Home Office.

Robinson, W. S. (1950). 'Ecological correlations and the behaviour of individuals.' *American Sociological Review* 15: 351-357.

The Stationery Office (1999). *Crime and Disorder Act*. London: Home Office.

Shepherd, J., Shapland, M., Scully, C. (1989). 'Recording by the police of violent offences; An accident and emergency department perspective.' *Medicine, Science and the Law*. 29: 251-257.

Smith, D. J. (2000). 'Changing situations and changing people', in A. Von Hirsch, D. Garland and A. Wakefield (eds.) *Ethical and social perspectives on situational crime prevention*. Oregon: Hart Publishing: 147-174.

Young, J. (2002). 'Crime and social exclusion' in M. Maguire, R. Morgan and R. Reiner (eds.). *The Oxford Handbook of Criminology*. Oxford: Oxford University Press: 457-490.

**PAPER 4**

**Corcoran, J., I.D Wilson and J.A. Ware**

“Predicting the Geo-Temporal Variations of Crime and Disorder.” Accepted for publication in the International Journal of Forecasting (Special Issue on Crime Forecasting)

A summary of this paper was presented at the *6<sup>th</sup> Annual Crime Mapping Conference* in Denver, USA 2002.

A poster summarising the work was presented at *Geovisualisation*, Albufira, Portugal 2002.

---

**Abstract**

*Forecasting the temporal geography of crime is a relatively new area of research, which is intended to facilitate the effective deployment of police resources. However, while often rich in spatial information, a database of crime incidents is usually sparse in terms of the number of crimes recorded for an area, often constraining the effectiveness of geo-temporal forecasting. Moreover, while traditional police boundaries provide a means by which the rate of crime can be measured they often fail to reflect the true distribution of criminal activity and thus do little to assist in the optimal allocation of police resources. Nonetheless, crime rates vary considerably between urban regions, and these disparities are often best accounted for by the change in use over time of urban locales by differing populations.*

*This paper introduces the first stage in the development of a computerised system designed to facilitate accurate crime incident forecasting by focusing upon geographical areas of concern that may well transcend traditional policing boundaries. The paper focuses upon the development of a practical solution, for use in an operational policing environment, which ameliorates the deficiencies of these rigid boundaries and moves towards a more dynamic methodology. The computerised procedure utilises a geographical crime incidence-scanning algorithm to identify clusters with relatively high levels of crime (hot-spots). These clusters provide sufficient data for training artificial neural networks (ANN) capable of modelling trends within them.*

**Keywords:** Crime Forecasting, Cluster Analysis, Geographical Information System (GIS), Artificial Neural Networks.

A system that intelligently interrogates a constantly updated database of crime incidence, providing indicators of where and when crime is likely to be highest would be of great utility in real-time police resource allocation. However, Gorr et. al. (2007) have shown that the predictive power of forecasting models is a product of the incidence count utilised and that these are generally low in relation to crime type, time and space, and subject to randomness.

This paper details a potential prediction framework for short-term, tactical deployment of police resources. The objective here is the identification of areas where the levels of crime are high enough to enable predictive models to be produced. This work differs from other recent studies dealing with hot-spot methods (for example, Ratcliffe and McCullagh 1999) and their statistical significance (for example, Chainey and Reid 2002). Whereas these researchers employ hot-spot methods as a means of visualising and comprehending crime distributions, here their utility is extended to use identified hot-spot regions as the foundation for predictive models.

The methodology presented in this paper follows three key stages (summarised in Figure 1). The first (spatial analysis) identifies geographical clusters; the second (cluster modelling) determines the data quality of each cluster; the third (prediction) develops a predictive model using the results of the previous stages.

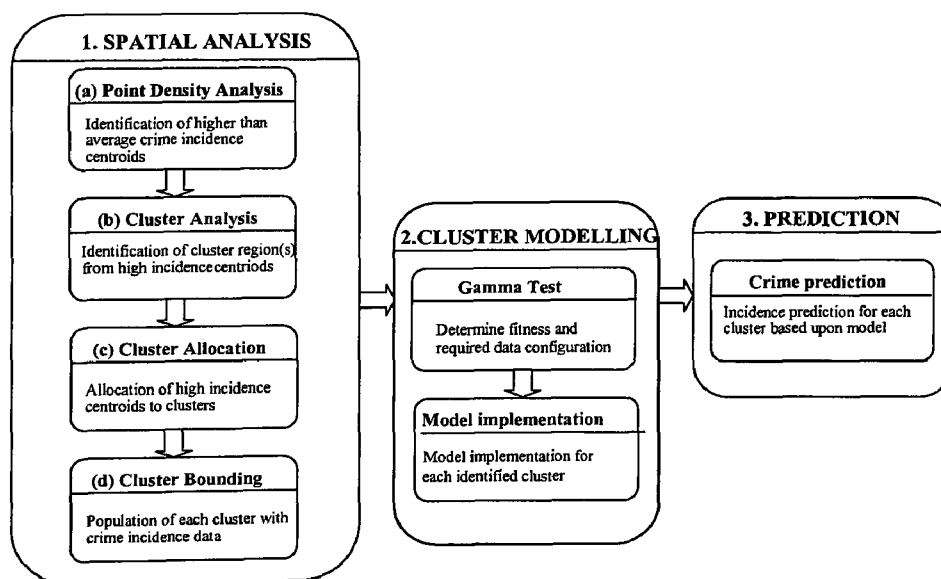
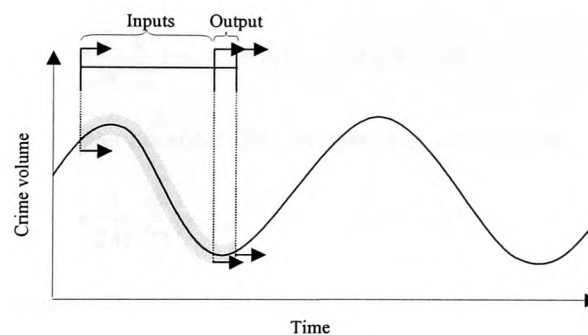


Figure 1 The CLAP/GT/Predictive process model

The paper demonstrates how a series of artificial neural networks can be trained, using geographical clusters of crime data, to facilitate predictive modelling (the extent to which each cluster has the potential to facilitate prediction is measured using a novel technique known as the Gamma Test (GT)). The paper also explains the details of the spatial analysis undertaken to geographically identify crime clusters. The paper concludes with a discussion of the results and focus for future research.

### Artificial Neural Networks Models for Prediction

Building forecasting models with neural networks is not a new phenomenon (see for example, Zhang, et. al. 1998 and Gorr 1994). In the case of crime level forecasting, the models tend to be autoregressive with input and output vectors being counts of crime. Here, the network model typically consists of multiple inputs  $y_{t-1} \dots y_{t-n}$  and a single output  $y_t$ . Training and test vectors are constructed by passing a moving window (illustrated in Figure 2) along the time series to extract its salient features.



**Figure 2 Moving window technique**

Testing the ANN involves presenting the network with a series of input vectors for which the tester, but not the network, knows the corresponding crime levels. The answers given by the network as to what it determines to be the level of crime, given the presented input vector, can then be used by the tester to determine the robustness of the training process. If the robustness of the network is deemed sufficient, the network can be used in a truly predictive capacity. Here the network is presented with an input vector for which the output is not known and its answer is assumed reliable.

## The Gamma Test (GT)

The GT (Stefánsson, 1997) is a non-linear data analysis technique that can be used to estimate the amount of noise inherent in a model constructed using a particular data set. The test estimates the best Mean Square Error (MSE) that can be achieved when modelling the data using any smoothing method, such as an ANN. The technique can be summarised by:

$$y = f(x) + \varepsilon \quad (1)$$

where  $y$  is the given output of an unknown function  $f$  and  $\varepsilon$  is that part of the output not account for by  $f$ .

Consider an input/output data set

$$\{(\mathbf{x}_i, y_i) \mid 1 \leq i \leq M\} \quad (2)$$

The GT estimates the variance of  $\varepsilon$  by generating lists,  $N[i, k]$ , of the  $k$ th ( $1 \leq k \leq p$ ) nearest neighbours  $\mathbf{x}_{N[i, k]}$  ( $1 \leq i \leq M$ ) for each vector  $\mathbf{x}_i$  ( $1 \leq i \leq M$ ). Specifically, the GT is derived from the input vector:

$$\delta_m(k) = \frac{1}{M} \sum_{i=1}^M |x_{N[i, k]} - x_i|^2 \quad (1 \leq k \leq p) \quad (3)$$

where  $|\dots|$  denotes Euclidean distance, and corresponding output value:

$$\gamma_m(k) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i, k]} - y_i)^2 \quad (1 \leq k \leq p) \quad (4)$$

where  $y_{N[i, k]}$  is the unique  $k$ th nearest neighbour of  $y_i$  in output space, which are found and fitted with the regression line:

$$\gamma = A\delta + \Gamma \quad (5)$$

of the points  $(\delta_m(k), \gamma_m(k))$  ( $1 \leq k \leq p$ ) (Durrant 2001). Here the  $p$  nearest neighbours (used by the GT procedure shown in Figure 3) is fixed and bounded ( $2 \leq p \leq M$ ).

---

```

Load input/output vectors  $\{(\mathbf{x}_i, y_i) : 1 \leq i \leq M\}$  where  $\mathbf{x}_i \in \mathbb{P}^m$  and  $y_i \in \mathbb{P}$ .
For  $i = 1$  to  $M$ 
  For  $j = 1$  to  $M$ 
     $InputDistance[i, j] = |\mathbf{x}_i - \mathbf{x}_j|$  ( $n$ -dimension Euclidean distance)
     $OutputDistance[i, j] = |y_i - y_j|$ 
  Next  $j$ 
Next  $i$ 
Sort corresponding  $InputDistance[i, j]$ ,  $OutputDistance[i, j]$  into ascending order.
For  $i = 1$  to  $M-1$ 
  For  $j = 1$  to  $M$ 
    If  $(InputDistance[i, j] = InputDistance[i+1, j])$  then

```

---



```

OutputDistance[i, j] = (OutputDistance[i, j] + OutputDistance[i+1, j]) / 2
InputDistance[i+1, j] = InputDistance[i+2, j]
Endif
Next j
Next i
For i = 1 to M
    For k = 1 to p
         $\delta(k) = \delta(k) + \text{InputDistance}[i, k]$  (Sum the input space distances.)
         $\gamma(k) = \gamma(k) + \text{OutputDistance}[i, k]$  (Sum the output space distances.)
    Next k
Next i
For k = 1 to p
     $\delta(k) = \delta(k) / M$  (Calculate the average.)
     $\gamma(k) = \gamma(k) / 2M$  (Calculate the half of the average.)
Next k
Compute linear regression on  $(\delta(k), \gamma(k))$  where  $(1 \leq k \leq p)$  as in (4).

```

Figure 3 The Gamma Test Algorithm

This algorithm produces an X,Y co-ordinate for each neighbour, which can be displayed using a two dimensional scatter graph. Figure 4 shows an example, where a regression line is plotted through the  $p$  nearest neighbours – the equation of the line,  $\gamma = A\delta + \Gamma$ , is also shown. Here, similar inputs plotted along the  $x$ , or Delta, axis will have similar output values plotted along the  $y$ , or Gamma, axis (see Evans and Jones (2002a, 2002b)).

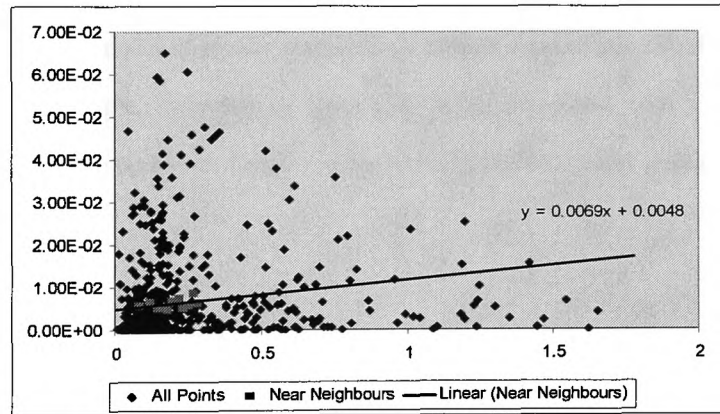
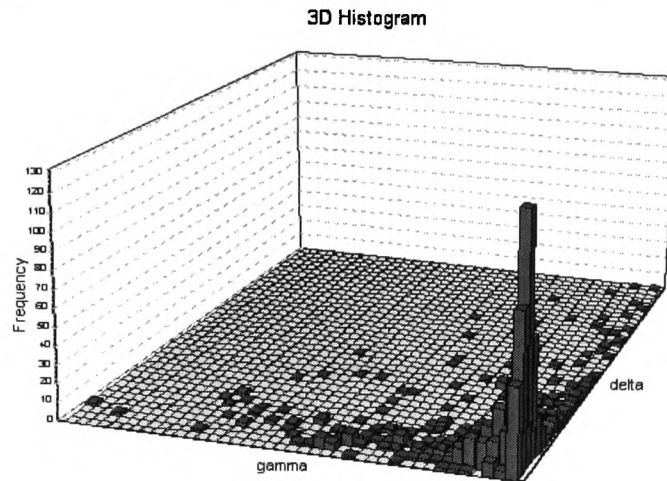


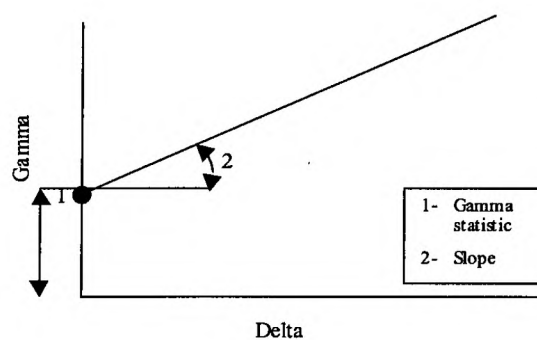
Figure 4 Gamma Test 2D graphical analysis

However, noise within the data will result in different Gamma plots for a given Delta value (illustrated using the frequency histogram shown in Figure 5). Ideally, the histogram should show a preponderance of points close to the origin, indicating that similar inputs are generally producing similar outputs. This provides a useful means for visualising patterns within numerical input/output data.



**Figure 5 Histogram of Delta, Gamma plot frequencies**

The graphical output, specifically the regression line (shown in Figure 6 without data points), provides two principal indicators. First, the vertical intercept  $\Gamma$  returned as the estimate for the variance of  $\varepsilon$ . This intercept on the  $y$  (or Gamma) axis offers an estimate of the best MSE achievable utilising a modelling technique, such as a neural network. Second, the gradient  $A$ , which offers an indication of model complexity (where a steeper gradient indicates a model of greater complexity). Results can indicate variations in the two variables (for example, estimates of low MSE being associated with a high level of complexity), with the preferred scenario being a low MSE and shallow gradient.



**Figure 6 The Gamma Statistic and the Gradient/slope**

Using the estimated MSE, a useful test is to establish the minimum quantity of data points needed to model an underlying function. This is accomplished using the M-Test, where the GT is applied to an increasing sample size ( $M_1 \dots M_m$ ) and the Gamma value plotted against that of  $M$ . In an ideal model the output may exhibit large variation in

Gamma at small values of  $M$  but ultimately stabilise at a higher value of  $M$ , which is indicative of the true noise variance inherent within the data. The region of asymptotic values of Gamma against  $M$  identifies the minimum data required to establish best possible accuracy in prediction.

The two indicators from the GT offer a basis from which an ANN model can be assembled and trained, given that an estimate of the best possible MSE has been provided by the GT. Hence, model training can be terminated once the estimated MSE is reached thus avoiding over-fitting. (For a more comprehensive discussion on the GT see Jones et. al. 2002.)

## The Crime Incident Data Set

The data that forms the basis of this study is comprised of 18,498 violent incidents (violence against the person, criminal damage and disorder), spanning one year in an urban area measuring approximately 242,700,000m<sup>2</sup>. Given this, it can be said that, on average, one crime took place during the year per 13,120m<sup>2</sup>, or approximately one crime per 65m radius. Included in the database of crime incidents are a number of variables relating to time, day, month, weather and location (represented as co-ordinates).

## Training Sets from Crime Clusters

For this analysis a similar technique to the GAM/1 geographical analysis machine developed by Openshaw (1987, 1988) was used, augmented to allow the clustering of centroids of high incidence. This four-stage process consists of:

- point density analysis;
- geographic representation and cluster analysis;
- allocation of centroids to clusters;
- finally, relation of incidents to cluster boundaries.

### *Stage One: Point density analysis*

During this initial stage, an analysis of the crime data, based on the hypothesis that areas with crime incidence greater than the average are significant, is carried out. The algorithm that looks for areas of greater than average crime incidence is shown in Figure 7 and the ensuing results for the test data set are illustrated in Figure 8.

---

*ScanRadius* is the radius of circle within which points will be counted.

```

StepSize is the distance that the centroid will be moved at each iteration.
StartX is the bottom left longitudinal start co-ordinate.
StartY is the bottom left latitudinal start co-ordinate.
EndX is the top right longitudinal end co-ordinate.
EndY is the top right latitudinal end co-ordinate.
Average is the number of crimes per area of circle.
For X = StartX to EndX step StepSize (moves longitudinally across the map)
    For Y = StartY to EndY step StepSize (moves latitudinally up the map)
        Count the number of points within ScanRadius
        If Count is greater than Average then
            Add co-ordinates to centroid list.
        Endif
    Next Y
Next X

```

Figure 7 Incidence density algorithm

Count is a function that iterates through the crime file, incrementing a counter each time one is found to be within *ScanRadius* of the centroid (*X*, *Y* co-ordinate), making use of Pythagoras:

$$\text{if}(\text{ScanRadius} < \sqrt{(\text{CrimeX} - X)^2 + (\text{CrimeY} - Y)^2}) \quad (6)$$

then add 1 to *count*

where *CrimeX* and *CrimeY* are projected co-ordinates of the crime incident.

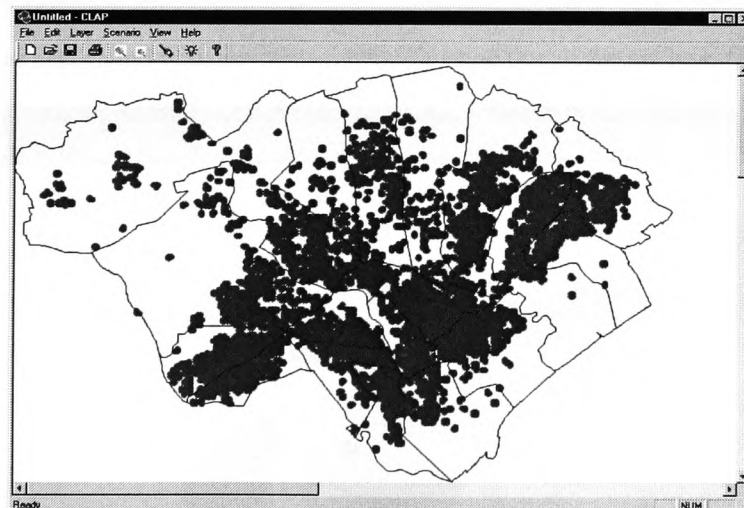
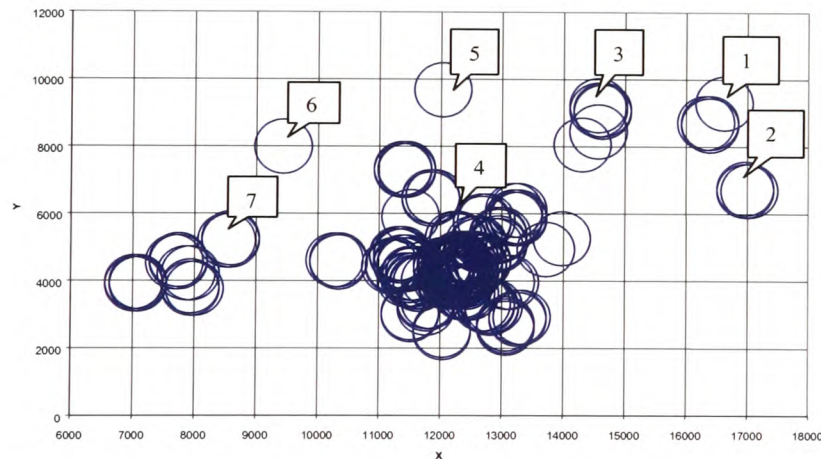


Figure 8 Centroids of higher than average crime incidence

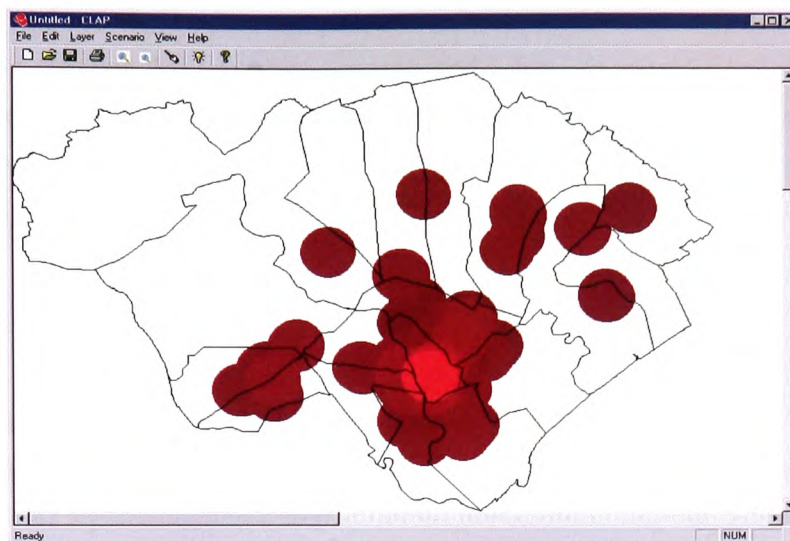
**Stage Two: Geographic representation and cluster analysis**

At this stage, a heuristic approach is taken to determine the level of crime incidence required for a cluster to be considered salient. The heuristic rules utilised to make this determination are based on an assumption that most incidence of crime tends to be concentrated within relatively small geographic areas.



**Figure 9 Clusters of high crime incidence**

Given this, a scatter graph representation of the geographical data was utilised to heuristically increase both the density of centroids displayed and the radius of the area associated with that centroid. In other words, as the density of the centroids displayed increases, so too does the radius of influence associated with that centroid. Experimentation resulted in the radius of influence, or *gravity*, being set to  $density \times 20$ , where *density* is the count of crimes associated with the centroid during stage one of the analysis process.



**Figure 10 Clusters of high crime incidence, displayed using the CLAP interface**

User interaction resulted in a centroid density of 40 being selected, resulting in the identification of seven clusters of interest (illustrated in Figure 9 and Figure 10). A list containing each of the salient centroids can now be produced.

**Stage Three: Allocation of centroids to clusters**

Next, the centroids which should be grouped together to form clusters are identified. The density and gravity parameters, together with the centroid list generated in stage two, forms the basis for this iterative procedure (outlined in Figure 11).

```

Enumerate each centroid and define each as being a potential cluster
While a cluster is growing
    For each cluster
        Attempt to Expand the cluster.
    Next Cluster
Wend
Remove duplicate clusters from list.

```

**Figure 11 Centroid clustering algorithm.**

*Expand* is a function that takes a cluster, in this case a list of centroids, which checks every other centroid to see if it *Intersects* with a member of the cluster. If a centroid is found that intersects with a member of the cluster that is not already a member of the cluster, it is added to the list.

*Intersects* is a function that takes two centroids and a distance value (*gravity* defined in stage two), which returns true if the distance between their centres is less than sum of the radii, according to the following equation:

$$\text{if } \left( (gravity * 2) < \sqrt{(c2.x - c1.x)^2 + (c2.y - c1.y)^2} \right) \text{ then} \quad (7)$$

*true else false*

where  $c_n.x$  and  $c_n.y$  are the central co-ordinates of each of the centroids.

**Stage Four: Relating incidents to cluster boundaries**

Finally, each of the clusters is populated with data ready for training a series of ANNs (one per cluster). The algorithm utilised to perform the population is outlined below:

```

For each cluster
    For each centroid within the cluster
        For each crime
            Add the crime to the cluster if the crime occurred within the centroid and
            the crime has not already been added and then add one to the count
            associated with this cluster.
        Next Crime
    Next Centroid
    Add Output results to unique cluster list
Next Cluster

```

**Figure 12 Incident to cluster boundary allocation algorithm.**

Equation (6) is again used to determine if a crime falls within the radius determined by the *gravity* value (800m). Each crime record contains a unique identifier, the cluster it belongs to and the weekday during which the crime was committed. In addition, each cluster record has a unique identifier, a list of its member centroids and a total crime count (shown in Table 1).

Cluster	Centroids	Crime Count
1	0, 4, 5, 6	1254
2	1, 2, 3	50
3	7, 8, 9, 10, 11, 12	954
4	13-109, 111-148, 165, 166	4097
5	110	161
6	149	228
7	150-164	2094
Total Clustered:		8838
Total Violent Crime:		18498

**Table 1 Crime incidence by cluster**

### Application of Gamma test to the cluster data

Taking the results from the cluster analysis two techniques were used to model the clustered data. The first sought to model day of week against crime volume, the second treated the data as a continuous times series using a windowing technique (Figure 2). Initial attempts using the first technique failed to achieve the accuracy accomplished by the second. The windowing technique was selected as the preferred procedure on the basis of the short-term forecasting problem. The following discusses the associated methodology.

### Forecasting using Artificial Neural Networks

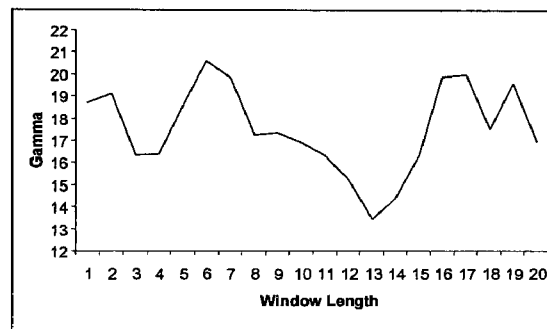
Implementation of an ANN model requires careful consideration of a series of model parameters each potentially impacting upon model stability and efficiency. These included decisions that concern architecture type, (number of input/output nodes and hidden layers), selection of training algorithm and volume of data to be used for training and testing.



### The network architecture

The ANN presented in this paper comprises of an input layer (corresponding to the length of the input vector), an output layer, providing the forecast value, and two layers of hidden nodes.

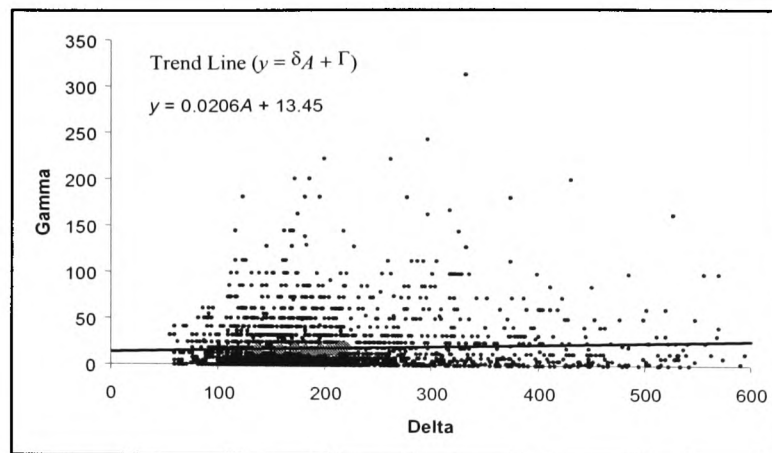
Modelling a time series involves generating a set of input vectors and corresponding output values. This is achieved by passing a moving window across the entire data set. For each movement of the window all values contained within it, apart from the last, form the input vector, while the last is the corresponding output value. Optimal window length for each cluster is established by generating a Gamma statistic for incremental window lengths. The optimal window length is that with Gamma statistic closest to zero (for example, 13 for the city centre cluster as shown in Figure 13).



**Figure 13 Example of increasing embedding (City Centre, cluster Four)**

The best topology for nodes in the hidden layers was determined empirically. Previous research has indicated that use of a single hidden layer is sufficient to learn any complex non-linear function (Hornik, 1991). However, Chester (1990), Srinivasan et. al. (1994) and Zhang (1994) suggest that two hidden layers can produce more efficient architectures. Initial large numbers of nodes ( $2N+1$  split evenly between both hidden layers, where  $N$  is the number of inputs) in the hidden layers were incrementally reduced to a minimum whilst maintaining acceptable forecasting capabilities. The shallow gradient (shown in Figure 14) suggested that a relatively few number of hidden nodes in proportion to the  $2N+1$  rule would be sufficient to model the underlying function and this proved to be the case.





**Figure 14 Gamma Scatter Plot showing the trend line**

A one step ahead forecasting horizon was used, represented by a single node in the output layer.

#### **Updating the weights**

The Conjugate ANN is an improved variation of the standard Backpropagation ANN. The standard gradient descent method for adjusting weights is replaced with Conjugate gradient descent (Bishop 1996), which uses past gradient measures to improve the error minimisation process.

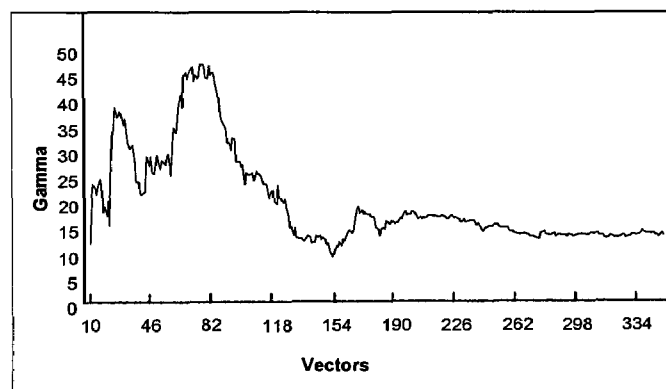
#### **Terminating the training procedure**

As over-fitting is a widely accepted problem associated with modelling utilising ANNs, the GT's ability to accurately measure the 'noise' within a data-set and, consequently, the point at which training should stop provides a significant utility for practitioners. Over-fitting occurs because the ANN will eventually attempt to fit all data encountered, including any noise present. Providing a measure of any noise present in the data set allows training to be terminated at a near optimal point. This is because an ANN will tend to fit useful data before any noise. Therefore, the GT statistic  $\Gamma$  provides a MSE value at which training can be stopped (for example, an approximate target value of 13.45, shown in Figure 14, for the City Centre cluster).

#### **Partitioning the vectors into training and test sets**

Once the number of inputs required to model the output is known the data can be transformed to fit the optimal set-up. Using this set-up, an M-test is performed to establish whether the available number of vectors is sufficient to model any underlying function. An

asymptotic level for the Gamma statistic (which approximates to the inherent noise of the output) indicates that there is sufficient data and provides a point where the data can be split into training and test vector sets. This is an important consideration as it allows the data set to be split into two rather than the commonly practised three parts (the training, validation and test set). Consequently, Wilson *et. al.* (2002) has demonstrated that there is no need to set aside part of the data as a validation set, which is used to determine when continuing to train would result in over-fitting allowing a higher proportion of data to be utilised during training. Thus, selection of the appropriate amount of data for modelling is confirmed at a point in advance of where the M-test reaches a stable level indicative of inherent noise (for example, an approximate volume of 300, shown in Figure 15, might be taken as the partitioning point).



**Figure 15 M-Test for the City Centre**

Alternatively, a non-asymptotic level indicates that either there is insufficient data or a useful underlying function can not be extrapolated from the data.

## Experimental Work

ANN and comparative linear regression forecasting models were constructed using the GT, and compared to a 'random walk' (RW) model. The RW forecasts the change from  $t_i$  to  $t_{i+1}$  based upon the average change from one period to the next. For example, taking the known number of crimes for a Thursday the forecast for the following day is based upon the average observed change (over the entire time series) between Thursday and Friday.

### Comparison between ANN, Linear Regression and Random Walk

As an example of the results obtained, the ANN models discussed here focus on two clusters analysed according to daily incident count.

### Cluster Seven (residential)

The GT was used to determine window length (30), training (308) and test (28) vector partitioning, and noise estimate (8.01, or 32% of the range, a high value indicating a very chaotic data series). In addition, the gradient statistic estimate (0.0159) suggested that relatively few hidden nodes (10 in each of the two hidden layers) would be required to reach the estimated Gamma statistic. The resultant ANN, Regression and RW models resulted in error percentages of approximately 31.3%, 30.5% and 30.9% of the range, respectively.

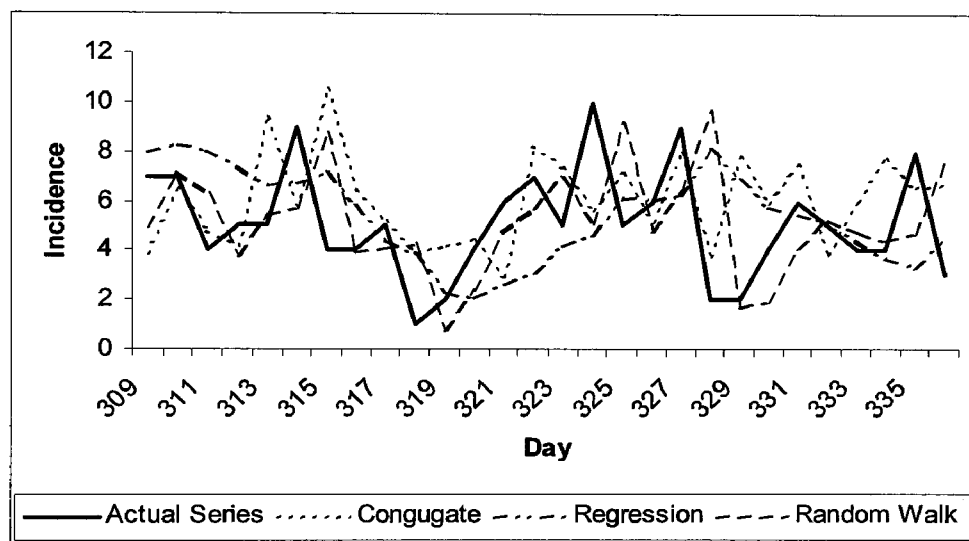


Figure 16 Incidence and forecast of violent crime (Cluster 7)

	Conjugate	Regression	Random Walk
Average MSE	7.95	7.61	7.77
Accuracy	31.3%	30.5%	30.9%
Min	1		
Max	10		
Range	9		

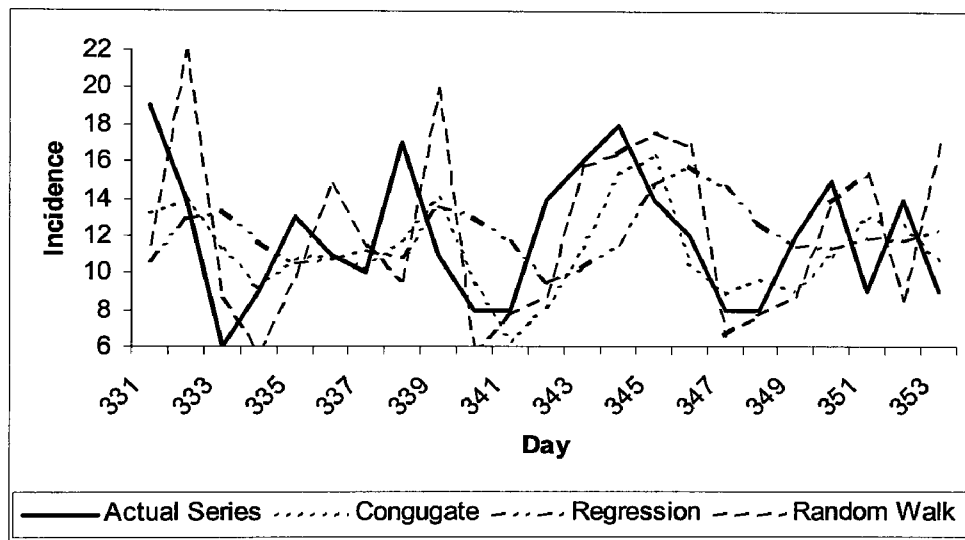
Table 2 Results for Cluster Seven

Cluster 7, which relates to a residential area with very few owner-occupiers, showed almost no general correlation between incident rate and weekday (illustrated in Figure 16).

However, an increased tendency for violent crime towards weekends was noted, warranting a closer examination of other causal factors.

### Cluster Four (City Centre)

The GT procedures were utilised to determine window length (13), training (330) and test (23) vector partitioning, and noise estimate (13.1365, or 27% of the range). In addition, the gradient statistic estimate (0.0159) suggested that relatively few hidden nodes (5 in each of the two hidden layers) would be required to reach the estimated Gamma statistic. The results generated by the resultant ANN, Regression and RW models (shown in Figure 17 and Table 3) produced error percentages of approximately 24.2%, 33.5% and 36.4% of the range, respectively.



**Figure 17 Incidence and forecast of violent crime (City Centre)**

	Conjugate	Regression	Random Walk
Average MSE	9.94	18.96	22.50
Accuracy	24.2%	33.5%	36.4%
Min	6		
Max	19		
Range	3		

**Table 3 Results for Cluster Four**

Cluster 4, which relates to a city centre, along with a concentrated collection of night clubs, public houses, public transport centre, and a sporting stadium, showed a higher incidence of crime at weekends, with peaks during times of known sporting events.

### **Discussion of results**

The results concurred with expectations given the GT's output. Importantly, exceptional incidence levels that occur infrequently, appear as noise, and are excluded from the underlying model accounting for a large portion of the error margin. The utility of the GT was demonstrated as a pre-model evaluation technique. Moreover, the results show that the methodology has the capacity to model the cause and effect relationship where one exists. The city centre (cluster four) offered the best predictive model using the ANN, cluster seven (residential area) generating relatively poor models for ANN, Regression and RW. Further experiments are now needed with an initial data set covering a longer period.

### **Conclusions and future work**

This paper introduces the first stage (summarised in Figure 1) in the development of a computerised system designed to facilitate accurate crime incident forecasting by focusing upon geographical areas of concern that may well transcend traditional policing boundaries. The paper focuses upon the development of a practical solution, for use in an operational policing environment, which ameliorates the deficiencies of these rigid boundaries and moves towards a more dynamic methodology. The computerised procedure utilises a geographical crime incidence scanning algorithm to identify clusters with relatively high levels of crime (hot-spots). These clusters provide sufficient data that can be analysed using the GT procedures assessing fitness and required configuration (for example, quantity of data required and number of inputs) for predictive modelling. Using the outputs from the GT two techniques were implemented (ANN and Regression) the ANN generally exhibiting a superior capacity to model the trends within each cluster. RW was used a naïve forecasting method, the results demonstrating a comparable forecasting accuracy to the other techniques for cluster seven (residential area) where the GT indicated a chaotic data series.

Future developments will include the modelling of more detailed scenarios to facilitate prediction based upon selected input criteria. Thus, for example the impact upon the region

of say a forthcoming public holiday, where the weather is predicted to be warm could be evaluated. The objective here was to extract an underlying, generalised, model of crime incidence. However, specific localities might best be modelled independently of the other data at specific times of the year (for example, sporting events that generate exceptionally high crime *spikes* that fall out of a generalised model). These spikes could be extracted and treated as a separate modelling exercise, given sufficient high quality data. Alternatively, a statistical analysis of exceptional events would provide an estimate of the change against normal levels that could subsequently be encoded as a set of rules that modify the incidence count accordingly. These two differing approaches will form the basis for continued experimentation.

## References

- Bishop, C. M., 1996, *Neural Networks for Pattern Recognition* (Oxford University Press).
- Chainey, S. and S. Reid, 2002, "When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime," In: D.B. Kidner et. al. (Eds.), *Socio-Economic Applications in Geographical Information Science*. Taylor and Francis.
- Chester, D. L., 1990, "Why two hidden layers are better than one?" *Proceedings of the International Joint Conference on Neural Networks*, 1, 265-268.
- Durrant, P. J., 2001, "winGamma™: A non-linear data analysis and modelling tool with applications to flood prediction," Ph.D. thesis, Department of Computer Science, University of Wales, Cardiff, Wales U.K.
- Evans, D. and A. J. Jones, 2002a, "A proof of the Gamma test", *Proc. Roy. Soc. Lond. A*, 458, 1-41.
- Evans, D. and A. J. Jones, 2002b, "Asymptotic moments of near neighbour distance
- Gorr, W., Olligschlaeger, A. and Y. Thompson, 200?, "Assessment of Crime Forecasting Accuracy for Deployment of Police," *International Journal of Forecasting, Special Issue on Crime Forecasting*.
- Gorr, W., 1994. "Editorial: Research prospective on neural network forecasting." *International Journal of Forecasting* 10(1): 1-4.
- Hornik, K., 1991, "Approximation capabilities of multi-layer feed-forward networks," *Neural Networks*, 4, 251-257.
- Jones, J. A., Evans, D., Margetts S. and P. J. Durrant, 2002, "The gamma test," *Heuristic and optimization for knowledge discovery*. R. Sarker, H. Abbass and C. Newton. (Hershey, Idea Group Publishing).
- Openshaw, S., 1987. "An automated geographical analysis system," *Environment and Planning A*. 19, 431-436.
- Openshaw, S., 1988. "Leukaemia patterns in Northern England: A new method of finding cancer clusters," *Northern Economic Review*. 16, 52-69.
- Phillips, D., 2000, "National Intelligence Model," (National Criminal Intelligence Service).
- Ratcliffe, J.H. and M. J. McCullagh, 1999, "Hotbeds of crime and the search for spatial accuracy," *Journal of Geographical Systems*, 1, 385-398.

Srinivasan, D. Liew, A. C. and C. S. Chang, 1994, "A neural network short-term load forecaster," *Electric Power Systems Research*, 28, 227-234.

Stefánsson, A., Koncar, N. and Antonia J. Jones, 1997, "A Note on the Gamma Test," *Neural Computing and Applications*, 5, 131-133.

Wilson, I. D., Paris. S. D., Ware, J.A. and D. H. Jenkins, 2002, "Residential Property Price Time Series Forecasting with Neural Networks," *Knowledge-Based-Systems*, 15, 335-341.

Zhang, G., B. Eddy Patuwo and M. Y. Hu., 1998. "Forecasting with artificial neural networks:: The state of the art." *International Journal of Forecasting* 14(1): 35-62.

Zhang, X., 1994, "Time series analysis and prediction by neural networks," *Optimisation Methods and Software*, 4, 151-170.



## PAPER 5

**Corcoran, J. and J.A. Ware.**

“Crime hot spot prediction: a framework for progress.” Socio-economic applications in geographical information science. Innovations in GIS 9. D. B. Kidner, G. Higgs and S. White. London, Taylor & Francis: (2003): 51-67.

A summary of this paper was presented at GISRUK 2001 and appears in the proceedings: Corcoran, J. and J. A. Ware (2001). Data Clustering using Artificial Neural Networks as a Precursor to Crime Hot Spot Prediction. GISRUK, University of Glamorgan, Wales: 249-253.

A summary of this paper was also presented at the 5<sup>th</sup> Annual Crime Mapping Conference in Dallas 2001

# Crime Hot Spot Prediction: A Framework for Progress

Jonathan Corcoran and J Andrew Ware

Artificial Intelligence Technologies Modelling Group, University of Glamorgan,  
Pontypridd, Mid Glamorgan, CF37 1DL, UK  
{jcorcora, jaware}@glam.ac.uk

**Abstract.** Crime rates differ between types of urban district, and these disparities are best explained by the variation in use of urban sites by differing populations. A database of violent incidents is rich in spatial information and studies have, to date, provided a statistical analysis of the variables within this data. However, a much richer survey can be undertaken by linking this database with other spatial databases, such as the Census of Population, weather and police databases. Coupling Geographical Information Systems (GIS) with Artificial Neural Networks (ANN) offers a means of uncovering hidden relationships and trends within these disparate databases. Therefore, this paper outlines the first stage in the development of such a system, designed to facilitate the prediction of crime hot spots. For this stage, a series of Kohonen Self-Organising Maps (KSOM) will be used to cluster the data in a way that should allow common features to be extracted.

## 1 Introduction

The advent of computers and the availability of desktop mapping software have advanced the analytical process, allowing efficient generation of virtual pin maps depicting crime incidents. A logical step beyond visualisation and analysis of trends and patterns is the development of a predictive system capable of forecasting changes to existing hot spots and the evolution of new ones.

Crime prediction in criminological research has been established for a number of decades (Ohlin and Duncan 1949; Glueck 1960; Francis 1971), although its foundation within a geographic and GIS context is still in its infancy.

As predictive crime analysis is a new research area, very little literature currently exists. Olligschlaeger (1997) provides an overview of existing forecasting techniques, concluding the time, level of user interaction and the expertise that each demands would be unrealistic for implementation in an operational policing environment. (Olligschlaeger and Gorr 1997) In addition, the inherent inflexibility and inability to dynamically adapt to change would compromise their viability in policing. ANN's are presented as one technique that offers minimal user interaction in addition to dynamic adaptability, and thus a potential operational forecasting solution.

### 1.1 Potential for Crime Prediction by the Police

A recent survey (Corcoran and Ware 2001) has highlighted the uptake and use of computer based crime-mapping technologies by British Police Forces. Computerised mapping technologies are rapidly becoming a vital prerequisite for visualisation of incident distributions and assisting in both the identification/allocation of resources and production/evaluation of policing strategies. The ability to efficiently generate simple maps to depict crime location and densities can be used directly to inform police officers and policing strategies, therefore maximising effectiveness and potential. A recent report published by the Home Office (Home-Office 2000) has underlined the importance of geographic data for analysis and interpretation of crime at the local level. In addition, the Chief Constable of Kent County Constabulary notes that "*over the last few years, police activity has shifted its centre of balance away from the reactive investigation after events, towards targeting active criminals on the balance of intelligence*". (Phillips 2000)

A natural step beyond visualisation and analysis of past and current incidents is the prediction of future occurrences, thus providing an *early warning system* for the Police.

(Olligschlaeger and Gorr 1997) Prediction can help prevent crime in that it facilitates the optimal allocation of limited resources. Prevention might be better than cure, but in the real world, very often, this is under financial constraints. The development of tools for prediction can thus help prevent crime and maintain optimal operational costs.

Prediction requirements for the police have been classified into three main categories according to the period of time involved (Gorr, Olligschlaeger *et al.* 2000):

- Short-term (tactical deployment);
- Medium-term (resource allocation);
- Long-term (strategic planning).

The focus for crime forecasting lies with short-term models as police tend to respond to existing and emerging crime patterns on relatively short time-scales for example on the basis of daily, weekly and monthly figures. (Gorr and Olligschlaeger 1998) This paper details a potential prediction framework for short-term, tactical deployment of police resources. The framework will allow identification of risk factors from which probabilities of criminal activity (specifically emergence of hot spots) can be derived and the necessary resources deployed.

## 1.2 COPDAT

The Crime and Offender Pattern Detection Analysis Technique (COPDAT), outlined in this paper, offers a potential framework that can be applied to geographic prediction. COPDAT entails the implementation of a GIS, integrating various spatial databases to analyse and map the identified trends.

## 2 Methodology

The volume of crime is insufficient to accurately predict an offence (in terms of location and time) when extrapolating from past offences. Therefore, in the proposed system, the type of crime predicted to take place within a particular time-window is supplemented by a separate prediction of the likely vulnerable areas for the same epoch. In addition, it would seem prudent to have the facility in a finished system to allow information based on police intelligence and experience to be built into the predictive model. The idea is to enhance current police predictive capabilities not replace them.

The spatial framework for the prototype COPDAT conforms to police sector boundaries for Cardiff (the capital city of Wales in the United Kingdom), whereby the volume of crime is sufficient to derive accurate predictions.

### 2.1 Data Sets

The accurate forecasting of the temporal-geography of crime (where and when a crime is likely to take place) would be of immense benefit, for accurate prediction if acted upon should lead to effective prevention. However, crime prediction frequently relies on the use of data appertaining to past perpetrators and/or past victims. Such data is therefore subject to legal and ethical restriction on its use, resulting in an ethical conundrum (Ware and Corcoran 2001). Therefore, the prototype COPDAT involves the use of only two central data sets - one pertaining to crime and the other providing contextual information.

### 2.2 GIS Techniques

Visual inspection of the various spatio-temporal data sets is a vital pre-requisite in assimilating an understanding of fundamental relationships and trends. The GIS is used as a tool to conduct this basic pattern analysis, including cluster and hot spot techniques.

### 2.3 Artificial Neural Network Techniques

ANN models provide a mechanism through which the various spatial, non-spatial and temporal data sets can be analysed to identify patterns and trends previously undiscovered. Identification and consolidation of such trends and patterns between the various data sets allows generalisation and subsequent prediction of future events and scenarios based upon that generality. For example, identifying that a specific kind of crime is likely to occur in a

certain location type under a given set of ambient conditions allows future incidents to be predicted.

The resultant of this scenario can be to produce a spatial probability matrix encapsulating a risk assessment of the study area. The spatial probability matrix can be directed to the GIS for visualisation in the form of a thematic contour map differentiating different areas with respect to their potential risk. In essence, the ultimate goal of the COPDAT is to learn decision criteria for assigning risk levels to new and future situations. This, for example, may involve identifying and predicting areas most at risk during a hot summer bank holiday football match in the City. Provision of such information is of obvious interest and of operational benefit to the police in terms of both resource allocation and policing strategies.

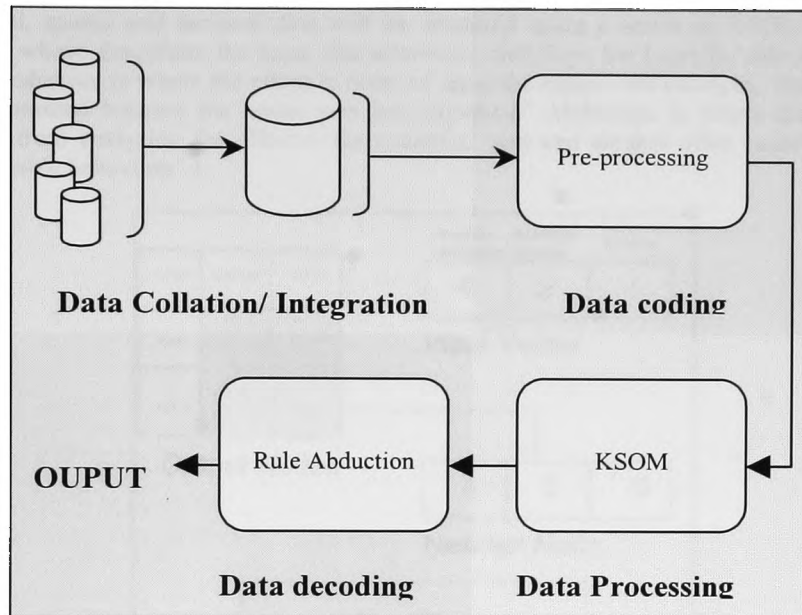


Figure 1 Data preparation process in relation to ANN processing

## 2.4 Data Preparation and Processing

Data representation and structuring is of key importance in the production of a robust predictive model. It has been shown in previous neural network studies that a certain level of pre-processing of the raw data is advantageous to model accuracy, efficiency and stability. This approach subsequently requires a certain level of post-processing in order to generate the required output values (illustrated in Figure 1).

## 2.5 Data Pre-processing

First, the data will undergo systematic and comprehensive analysis. This is followed, where necessary, by converting the data into an alternative representation more suited for input into an ANN. This sequential process can be broken down to a series of discrete stages:

- Format conversion and integration;
- Error detection and editing;
- Data reduction, transformation and generalisation.

The final stage in the pre-processing is of critical consequence in terms of a successful ANN implementation. The process of feature extraction and encoding of such characteristics impacts upon the ANN's ability to learn and assimilate relationships between salient variables and hence its accuracy in prediction. This process can be further decomposed into three distinct procedures:

1. Transformation and scaling may include:
  - Applying a mathematical function (e.g. logarithm or square) to an input;

- Scaling the different inputs so that they are all within a fixed range can greatly effect the reliability of an ANN system.
2. Reduction of relevant data includes simple operations such as filtering or taking combinations of inputs to optimise the information content of the data. This is particularly important when the data is noisy or contains irrelevant and potentially erroneous information.
  3. Encoding of identified features for input to the ANN. The data types include a range of data categories (discrete, continuous, categorical and symbolic), each to be handled and represented in an explicit manner.

## 2.6 Clustering Using a Kohonen Self Organising Map

Temporal, spatial and incident data will be clustered using a series of KSOM. These clusters, whose data share the same characteristics, will form the basis for rule abduction. (Note, deduction is where the effect is deduced from the cause - for example, 'the burglary was committed because the house was left unlocked.' Abduction is where the cause is gleaned from analysing the effect - for example, 'sun and alcohol often causes socially unacceptable behaviour'.)

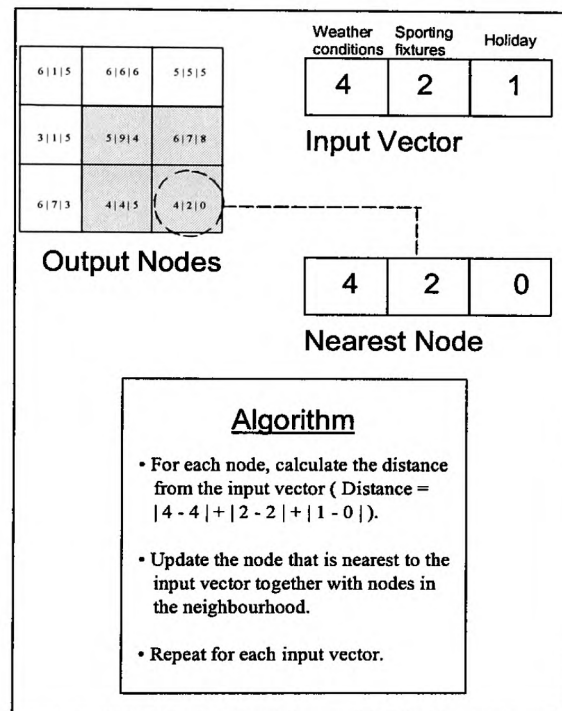


Figure 2 A Kohonen Self Organising Feature Map.

An unsupervised network, such as the Kohonen Self Organising Map (KSOM), organises itself in such a way as to represent classes within a data set. The 2-D KSOM allows classes to be visualised on a feature map, in which similar inputs are spatially clustered. Figure 2 shows a typical 2D KSOM along with an abridged algorithm (Note, the number of nodes are arbitrarily selected for example purposes).

Each output node on the KSOM contains a vector of length 'j', where 'j' is equal to the number of input attributes. Before training begins, the network is placed into an initialised state, i.e. the directions of the vectors in each node are randomised. Training involves passing an input vector into the network through the input nodes. Each node on the KSOM is then compared with the input vector, and the closest node is then changed to be more like the input vector. Neighbouring nodes also become more like the input vector. Iterating this process achieves clustering of similar input vectors in Euclidean space.

## 2.7 An Overview of the Methodology

The methodology uses a KSOM to find clusters in the input vectors and then the data from each cluster is used to train a separate MLP network. The advantage of using the KSOM for this application is that it can identify clusters within the parent dataset that are difficult to identify using simple sort procedures. Figure 3 gives an overview of the method. A dataset containing the required elements of the vector  $v$  is passed through the KSOM during the training stage and allowed to develop into clusters. After training, the clusters are inspected and the primary clustered features used to describe the sub-datasets. These sub-datasets are then used as the basis for rule abduction.

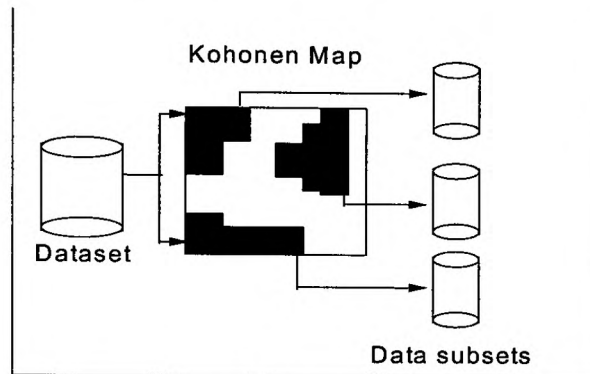


Figure 3 Using a KSOM to partition a dataset

However, two fundamental problems need to be resolved before this method can be of any use. First, the problem of separating adjacent clusters, and second, the desire to proceed to the abduction phase only using 'good' clusters (see Figure 6).

The first problem has been recognised in other studies and some guidelines have been provided. (James 1994) In essence, the problem lies in the attribution of boundary nodes to a specific cluster. Figure 4 provides an example of a KSOM output with adjacent clusters. There appear to be four classes within the dataset, but there are regions of uncertainty relating to the boundaries of each cluster. (The Digits in each node show the number of vectors mapped to that node.)

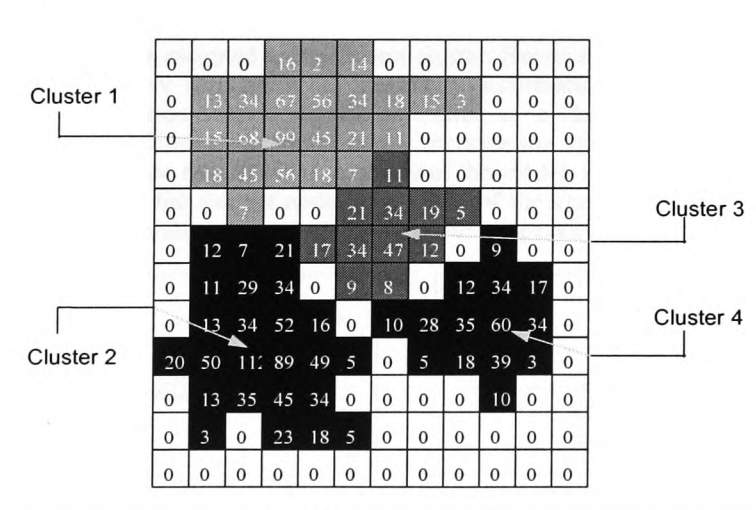


Figure 4 An Example Trained Kohonen Self Organising Feature Map

To overcome this problem a simple method of identifying class boundaries or discriminants can be used, which relies on the fact that a KSOM clusters primarily on binary features. For example, if the type of crime is represented using binary inputs, the KSOM will tend to cluster records according to this attribute. Boundaries between adjacent clusters on a 2D map can then be found by inspecting the records mapped to each node and grouping

together nodes that contain the same classification values. However, this level of clustering can be achieved using a multi-level sort procedure. In essence, the binary representation of the data will dictate the make-up of the resulting clusters and more importantly the homogeneity of the data sets.

If the data are represented using continuous inputs, the clusters formed by the KSOM would provide more generalised classes which would be difficult to achieve using a sort procedure. However, the inspection method would no longer identify class boundaries, as the similarities between records would not always be apparent. Clearly, before meaningful training data sets can be formed, the problem of discerning effective class boundaries in a KSOM must be addressed. Ideally, the network adaption rule should cluster similar inputs and clearly distance individual clusters. Zurada (1992) explains "One possible network adaption rule is: A pattern added to the cluster has to be closer to the centre of the cluster than to the centre of any other cluster". Using this rule, each node can be examined and the distance from the surrounding centroids can be calculated. (A centroid is taken to be a node, outside any known cluster boundaries, that has the largest number of input vectors mapped to it.) The subject node can then be added to the nearest cluster. Figure 5 illustrates a hypothetical situation where it is unclear where to draw the boundaries around clusters on a KSOM.

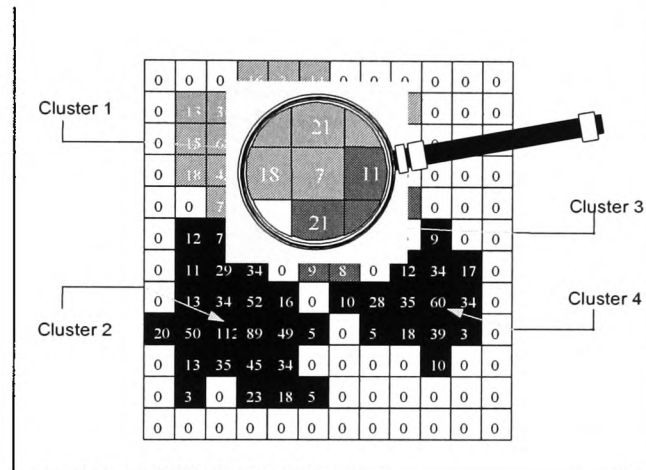
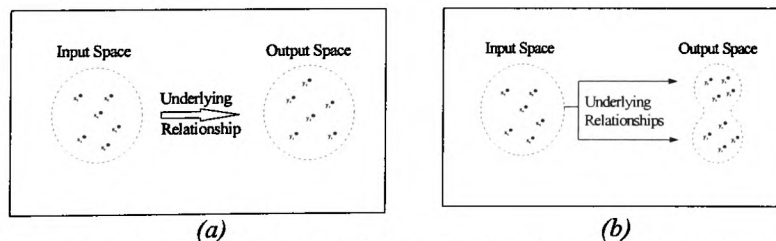


Figure 5 An Example KSOM.

By simply calculating the Euclidean distance of the subject node from the two centroids, the subject node can be assigned to the closest cluster. However, in this application, which aims to generate clusters with latent but meaningful information that can be subsequently extracted using abduction techniques, the formation of a class boundary for Cluster 1 (including the subject node) may dramatically increase the variance of the training data. This increase will reduce the potential accuracy of the model. In the example, it may have been better to exclude the subject node from either of the clusters, and deem the vectors mapped to the subject node as either being outliers or a separate cluster.



(a) An Example of a Good Input Cluster. A one-to-one relationship can be established and hence Input Space is homogeneous

(b) An Example of a Bad Input Cluster. Two or more similar vectors in the Input Space map to different vectors in the Output Space. Hence, the Input Space is not homogeneous.

Figure 6 Example cluster mappings from input to output space

In addition to identifying boundaries around input clusters, it is also important for this application to match input clusters to appropriate output clusters. In terms of criminal activity, if, for example, the KSOM has clustered crimes from two different locational areas, it is reasonable to expect these crimes to have similar characteristics.

Figure 6(a) illustrates a cluster of similar input vectors. When the corresponding data in output space is examined, all the examples describe similar output values. Conversely, Figure 6(b) shows a situation where the data can only be modelled using two or more functions. The problem now is to estimate the 'usefulness' of a given cluster. There are a number of options available of which the following are the most useful:

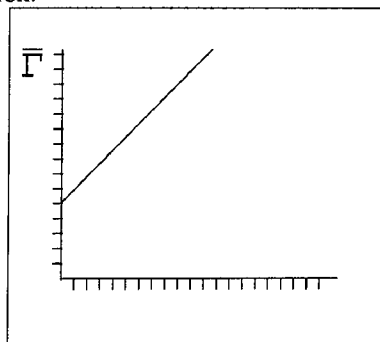
- Multi-Layered Perceptron (MLP) Model (Chen 1997)
- Class Entropy (Quinlan 1986)
- $R^2$  Almy (Almy 1998)
- Gamma Test (Stefánsson 1997)

For classification problems, Class Entropy can be used to decide if input clusters are homogenous with respect to output clusters. For example, Quinlan's C4.5 and C5.0 (Quinlan 1993) uses Class Entropy to segment the input space until each segment points toward a single class in output space. However, this approach is not applicable for regression problems such as this one and this rules out the use of class entropy.

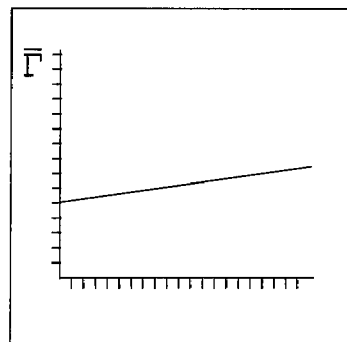
A quick and easy estimate of the susceptibility of a dataset for function induction can be achieved by executing a multiple regression analysis on the data and use the  $R^2$  value to discern trainable clusters. This technique is useful for data where the function is known to be linear. However, this is not known to be true for crime analysis data.

**The Gamma Test.** The Gamma test attempts to estimate the best mean square error that can be achieved by any smooth modelling technique using the data. If  $y$  is the output of a function then the Gamma test estimates the variance of the part of  $y$  that cannot be accounted for by a smooth (differentiable) functional transformation. Thus if  $y = f(x) + r$ , where the function  $f$  is unknown and  $r$  is statistical noise, the Gamma test estimates  $\text{Var}(r)$ .

$\text{Var}(r)$  provides a lower bound for the mean squared error of the output  $y$ , beyond which additional training is of no significant use. Therefore, knowing  $\text{Var}(r)$  for a data set allows prediction beforehand of what the MSE of the best possible neural network trained on that data would be. The Gamma test provides a method of determining the quality of the data stratification - a good stratification technique will result in a low value of  $\text{Var}(r)$  for each subset. Interpreting the output from the Gamma test requires considerable care and attention.



(a) High noise (large  $\bar{\Gamma}$  value) and high complexity (steep gradient)



(b) High noise (large  $\bar{\Gamma}$  value) and low complexity (flat gradient)



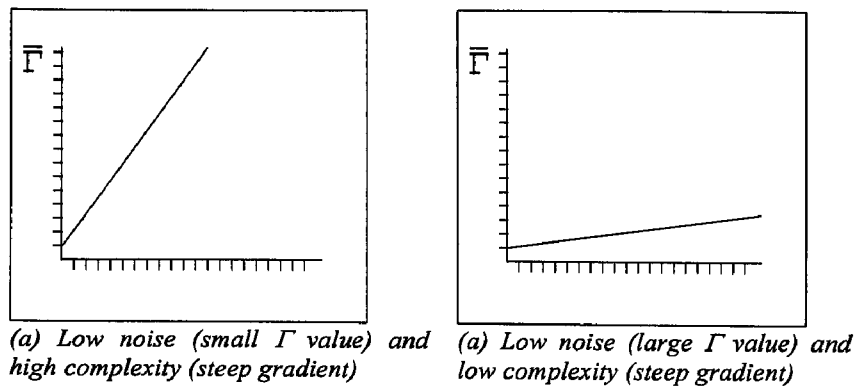


Figure 7 Interpreting the output from the Gamma Test

The least squares regression line provides two pieces of information. First, the intercept on the Gamma axis is an estimate of the best MSE achievable by any smooth modelling technique. Second, the gradient gives an indication of the complexity of the underlying smooth function running through the data.

The Gamma test may estimate a very low MSE but unfortunately show a high level of complexity that could be difficult to accurately model. It is easier to see this situation when the output from the Gamma test is presented graphically.

A hypothetical example with high noise content and high complexity is shown in Figure 7(a); high noise and low complexity Figure 7(b); low noise and high complexity in Figure 7(c); and, finally, low noise and low complexity (the desired outcome) in Figure 7(d). In summary, for this methodology to be successful, the following is required:

- class boundaries must be identified around clusters formed by the KSOM over the input space that exclude outliers and nodes from neighbouring clusters, and;
- only 'good' clusters (illustrated in Figure 6) should go on to form training data sets for subsequent back propagation models.

**A Detailed Look at the Methodology.** The Gamma test can be used at a number of abstraction levels within the KSOM stratification method:

- Cluster level;
- Node Level;
- Record Level.

Data stratification is achieved at cluster level or at node level, depending on the ease at which cluster boundaries can be determined. The record level gives an indication of outliers.

*Cluster Level Analysis.* This can be achieved thus:

*Identify Cluster boundaries in KSOM*

*For every cluster*

*Place records mapped to cluster into a file*

*Apply Gamma test to data in the file*

*If  $Var(r) \leq \text{some Threshold}$  then*

*Use data file as the training set for a MLP*

*else*

*Process at Node Level*

*Endif*

*Next Cluster*

This level of abstraction is the least computationally intensive as it only requires one pass of the Gamma test for each cluster. The disadvantage with this method is that it is often difficult to identify boundaries manually. In this case the Gamma test should be applied at the Node level.

*Node Level Analysis.* At this abstraction level, the methodology attempts to identify useful clusters by selecting a centroid and adding neighbouring nodes - where the addition of a

node increases the variance significantly it is subsequently removed. This process iterates until the cluster size is maximised within a specified variance threshold. This algorithm identifies useful clusters on a 2D KSOM. This is achieved thus:

```

number_of_clusters:=0
While there are nodes to cluster
    number_of_clusters := number_of_clusters + 1
    Select the unclustered node with the largest record count
    Apply Gamma test to the data in the selected node
    If Var(r) <= Threshold then
        nodes_of_interest := None
        (cluster includes only the data from selected node)
        For each unclustered node next to selected node
            Add data from unclustered node to the cluster
            Run gamma test on cluster
            If Var(r) <= Threshold then    Add node number to nodes_of_interest
            else Remove data from the unclustered node from the cluster
        While nodes_of_interest <> None
            Select c_node from nodes_of_interest
            Remove c_node from nodes_of_interest
            For each unclustered node immediately surrounding c_node
                Add data from unclustered node to the cluster
                Run gamma test on cluster
                If Var(r) <= Threshold then    Add node to nodes_of_interest
            else
                Remove data from the node from cluster
                Record the boundaries of this cluster
    else Process at Record Level

```

The boundary detection algorithm for a 1D KSOM is very similar except neighbouring nodes are selected progressively further away from the left and the right of the centroid node. This level of analysis is more computationally intensive the cluster level analysis, as it requires  $m \times \sum n_i$  passes of the Gamma test, where 'i' is the number of nodes investigated for cluster 'n' for a KSOM containing 'm' clusters. If using either the cluster level analysis or the node level analysis, useful clusters have been identified, it is then possible to train an independent MLP on each subset. The KSOM is then used to select the appropriate MLP on which to predict the value of a previously unseen example. The resulting system is closely related to a panel judgement system. However, if both methods have still resulted in poor training sets (useless clusters) then the analysis is taken to the most detailed abstraction level, that is the record level.

**Record Level Analysis.** The record level analysis is the most computationally intensive. The purpose of this level of the methodology is to identify data subsets from examples that have mapped to the same node on the KSOM. This facilitates extraction of outliers from a data set as well as giving some indication as to the examples that require additional features. The algorithm developed for this level of analysis is very similar to that shown for the node level analysis. However now, sets of records are iteratively analysed using the Gamma test. This is achieved thus:

```

For each node in the KSOM
    Apply Gamma test to estimate the variance for the data in node
    If Var(r) > Threshold then
        For each record at node
            Remove record from data set
            Apply Gamma test to estimate the variance for the data in node
            If New Var(r) < Previous Var(r) then Mark record as outlier
            else Add record back into data set
    else Proceed at Node Level

```

This level of analysis will identify the need for additional features and highlight records that may be classed as outliers.

## 2.8 Post-processing

Natural language rules will be derived from each 'good' cluster found by the KSOM. Each rule will represent a broad and generic definition (that with time can be fine tuned) of a specific sub-model that can be applied to best predict crime, for example:

*for Centre of City  
if Weather includes Wet  
and Day is Friday  
and Time is Night  
then Problems will include inside pubs (probability 0.9)*

*for Centre of City  
if Weather includes Dry  
and Day is Friday  
and Time is Night  
then Problems will include inside pubs (probability 0.4)*

*for Centre of City  
if Weather includes Dry  
and Weather includes Warm  
and Day is Friday  
and Time is Night  
then Problems areas will include outside pubs (probability 0.9)*

These rules together with other determining factors (including temporal information such as the day, time and prevailing weather) are then directed to the GIS for production of a thematic crime risk contour map. When used in a predictive manner, the GIS can provide an important visual reference for analysing the relative impact of multiple factors on crime levels in a given area.

## 3 Implementation Details

A series of KSOMs will provide the mechanism for classifying the heterogeneous nature of crime and criminal activity in a spatial, temporal and contextual framework. The assumption that there is a single simple linear relationship between the various factors is unrealistic and misplaced. In the light of this, a valid approach to assimilate an understanding of crime and criminal activity is to dissemble it into its homogeneous components for independent analysis. This analysis can be used to generate a series of rules and generalisations that can be utilised in a predictive capacity at a broader heterogeneous level.

The KSOM provides an unsupervised technique, through which this stratification process can be automated, resulting in the formulation of homogeneous clusters (illustrated in Figure 8). Where clusters are not clearly delineated it may be necessary to calculate a function such as the Euclidean distance or Gamma test (Lewis, Ware *et al.* 1997b, 1997a) in order to assign output to specific clusters.

The user will be able to present the trained suite of KSOMs, in the COPDAT, with structured scenarios from which probabilities and vulnerabilities of criminal activity will be derived and presented to the user via spatial thematic representations. Using standard GIS functionality, the user could subsequently overlay additional information, which may for example involve identification of specific localities, (street names, building etc) for deployment of resources.

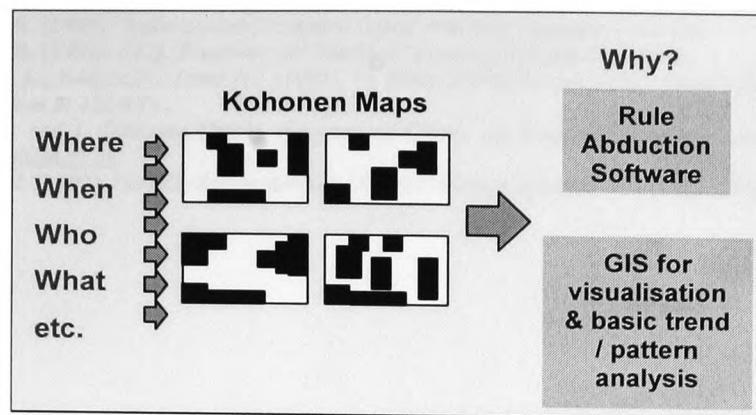


Figure 8 Proposed COPDAT Hybrid System

#### 4 Summary & Benefits

Coupling the predictive capabilities provided by the proposed COPDAT model with existing GIS techniques would enable law enforcement agencies to more effectively evaluate resource and tactical policies. Crime mapping and analysis using COPDAT will potentially facilitate research, assist with offender management and monitoring, and enable community planners to develop policy and forecast future needs for public safety resources. Provision of such facilities will rely upon use of multiple data sources and be dependent upon the use of pre-defined data processing and encoding techniques that will require new standards and techniques to be integrated into current practices. The Chief Constable of Kent County Constabulary raises this point in respect to police intelligence noting that "*this process is only possible if we can mix and match information across the board. To this end, it is essential that common standards and discipline attach to the intelligence process.*" (Phillips 2000)

This paper outlines a potential framework that can be expanded to model and predict a broader variety of crimes based upon a larger mixture of criminal and contextual data sets and a common schema for crime prediction.

#### References

- Almy, R., Horbas J., Cusack, M., Gloudemans, R., Ed. (1998). *The Valuation of Residential Property using Regression Analysis*. Computer Assisted Mass Appraisal: An International Review, Ashgate Publishing Company, England
- Chen, K., Xiang, Y., Huisheng, C. (1997). "Combining Linear Discriminant Functions with Neural Networks for Supervised Learning." *Journal of Neural Computing and Applications* 6.
- Corcoran, J. and J. A. Ware (2001). Application of Computer Crime Mapping Technologies by British Police Forces. Available from: jcorcora@glam.ac.uk.
- Francis, S. H. (1971). Prediction methods in criminology, *Home Office Research Studies*.
- Glueck, S. E. (1960). *Predicting delinquency and crime*. Cambridge, Massachusetts, Harvard University Press.
- Gorr, W., A. Olligschlaeger and Y. Thompson (2000). "Assessment of Crime Forecasting Accuracy for Deployment of Police." *International Journal of Forecasting*.
- Gorr, W. and A. Olligschlaeger (1998). *Crime Hot Spot Forecasting: Modelling and Comparative Evaluation*. <http://www.ojp.usdoj.gov/cmrc/grants/welcome.html>.
- Home-Office (2000). *Review of Crime Statistics: A Discussion Document*. <http://www.homeoffice.gov.uk/crimprev/crimstco.htm>.
- James, H. (1994). *An 'Automatic Pilot' for surveyors*, RICS Cutting Edge.
- Lewis, O. M., J. A. Ware and D. H. Jenkins, Eds. (1997a). *A Novel Neural Network Technique for Modelling Data Containing Multiple Functions*. Computational Intelligence - Theory and Applications, Springer Verlag 141-149.
- Lewis, O. M., J. A. Ware and D. H. Jenkins (1997b). "A Novel Neural Network Technique for the Valuation of Residential Properties." *Journal of Neural Computing and Applications* 5: 224-229.
- Ohlin, L. E. and O. D. Duncan (1949). "The efficiency of prediction in criminology." *American Journal of Sociology*: LIV 5.
- Olligschlaeger, A. and W. Gorr (1997). *Spatio-Temporal Forecasting of Crime: Application of Classical and Neural Network Methods*. <http://www.heinz.cmu.edu/wpapers/active/wp00162.html>.
- Phillips, D. (2000). *National Intelligence Model*, National Criminal Intelligence Service.

- Quinlan, J. R. (1986). "Induction of Decisions Trees." *Machine Learning* 1: 81-106.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*, Morgan Kaufmann.
- Stefánsson, A., Koncar, N., Jones, A. J. (1997). "A Note of the Gamma Test." *Neural Computing and Applications* 5: 131-133.
- Ware, J. A. and J. Corcoran (2001). *Forecasting Crime: An Ethical Conundrum*. Available from: [jaware@glam.ac.uk](mailto:jaware@glam.ac.uk)
- Zurada, J. M. (1992). *Introduction to Artificial Neural Network Systems*, West Publishing Company.